AF/1646

Corres. and Mail
BOX AF

Docket No.: PF-0040 US
Response Under 37 C.F.R. 1.116 - Expedited Procedure
Examining Group 1646

Certificate of Mailing
I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Box AF, Commissioner for Patents, Washington, D.C. 20231 on October 2, 2002.

By: Debbie Ellis Printed: Debbie Ellis

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Coleman et al.

Title: A C5a-LIKE SEVEN TRANSMEMBRANE RECEPTOR

Serial No.: 08/462,355 Filing Date: June 5, 1995

Examiner: Ulm, J. Group Art Unit: 1646

BOX AF
Commissioner for Patents
Washington, D.C. 20231

TRANSMITTAL FEE SHEET

Sir:

Transmitted herewith are the following for the above-identified application:

1. Return Receipt Postcard;
2. Brief on Appeal, including Appendix (24 pp., in triplicate); and
3. Three (3) references (in triplicate).

The fee has been calculated as shown below.

X	Fee for filing a Brief in support of an Appeal under 37 CFR 1.17(c):	$ 320.00
X	Please charge Deposit Account No. **09-0108** in the amount of :	$ 320.00

The Commissioner is hereby authorized to charge any additional fees required under 37 CFR 1.16 and 1.17, or credit overpayment to Deposit Account No. 09-0108. **A duplicate copy of this sheet is enclosed.**

Respectfully submitted,

INCYTE GENOMICS, INC.

Date: October 2, 2002

Susan K. Sather
Reg. No. 44,316
Direct Dial Telephone: (650) 845-4646

Customer No.: 27904
3160 Porter Drive
Palo Alto, California 94304
Phone: (650) 855-0555
Fax: (650) 845-4166

#23/gpw
10-16-02

Docket No.: PF-0040 US

Response Under 37 C.F.R. 1.116 - Expedited Procedure
Examining Group 1646

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Box AF, Commissioner for Patents, Washington, D.C. 20231 on October 2, 2002.

By: _____ Printed: Debbie Ellis

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Application of: Coleman et al.

Title: A C5a-LIKE SEVEN TRANSMEMBRANE RECEPTOR

Serial No.:	08/462,355	Filing Date:	June 05, 1995
Examiner:	Ulm, J.	Group Art Unit:	1646

Box AF
Commissioner for Patents
Washington, D.C. 20231

BRIEF ON APPEAL

Sir:

Further to the Notice of Appeal filed July 25, 2002, and received by the USPTO on August 2, 2002, herewith are three copies of Appellants' Brief on Appeal. Authorized fees include the $ **320.00** fee for the filing of this Brief.

This is an appeal from the decision of the Examiner finally rejecting Claims 12-17 of the above-identified application.

(1) REAL PARTY IN INTEREST

The above-identified application is assigned of record to Incyte Pharmaceuticals, Inc. (now Incyte Genomics, Inc.) (Reel 7659, Frame 0629) which is the real party in interest herein.

(2) RELATED APPEALS AND INTERFERENCES

Appellants, their legal representative and the assignee are not aware of any related appeals or interferences which will directly affect or be directly affected by or have a bearing on the Board's decision in the instant appeal.

(3) STATUS OF THE CLAIMS

Claims rejected:	Claims 12-17
Claims allowed:	(none)
Claims canceled:	Claims 1-11
Claims withdrawn:	Claims 18-22
Claims on Appeal:	Claims 12-17 (A copy of the claims on appeal, as amended, can be found in the attached Appendix).

(4) STATUS OF AMENDMENTS AFTER FINAL

There were no amendments submitted after Final Rejection.

(5) SUMMARY OF THE INVENTION

Appellants' invention is directed, *inter alia*, to polynucleotides encoding polypeptides having strong homology to canine C5a anaphylatoxin receptor (Perret, J.J., et al., (1992) Biochem. J., 288:911-917) ("CALR") and compositions containing them, which have a variety of utilities, in the diagnosis of conditions or diseases characterized by expression of CALR and for drug discovery (see the Specification at, e.g., page 6, line 15 through page 7, line 1; page 9, lines 13-22). As described in the Specification:

> The novel C5a-like receptor (CALR) which is the subject of this patent application was identified among the cDNAs derived from a mast cell library. Incyte Clone No. 8118 is a novel nucleotide sequence which is more closely related to CFCOMC5AM, the C5a anaphylatoxin receptor from dog (Perret JJ et al (1992) Biochem J 288:911-17) than to the known human C5a receptor. (Specification, page 2, lines 8-12.)

* *

The present invention provides a unique nucleotide sequence identifying a novel C5a-like receptor which was first identified in human mast cells. The sequence for calr is shown in SEQ ID No 1 and is homologous to the GenBank sequence, CFCOMC5AM for canine C5a anaphylatoxin receptor. Incyte 8118 has 45% amino acid identity with the C5a receptor and differs from it in having only three carboxylate residues in the N-terminus, two of which are Glu rather than Asp. In addition, the N-terminus of Incyte 8118 is shorter than that of the published C5a receptor and would be expected to have different binding specificity.

Because CALR is specifically expressed in cells active in immunity, the nucleic acid (calr), polypeptide (CALR) and antibodies to CALR are useful in investigations of and interventions in the normal and abnormal physiologic and pathologic processes which comprise the mast cell's role in immunity. Therefore, an assay for upregulated expression of CALR can accelerate diagnosis and proper treatment of conditions caused by abnormal signal transduction due to anaphylactic or hypersensitive responses, systemic and local infections, traumatic and other tissue damage, hereditary or environmental diseases associated with hypertension, carcinomas, and other physiologic or pathologic problems. (Specification, page 6, lines 7-22.)

* *

The cDNA (SEQ ID NO 1) and amino acid (SEQ ID NO 2) sequences for human CALR are shown in Fig 1. Incyte's calr produced a BLAST score of 412 when compared with the C5a receptor sequence and has a probability of 1.8^{-50} that the sequence similarity occured by chance. This calr homolog also resembles various N-formylpeptide receptors generating BLAST scores ranging from 381 to 363 with probabilities of 7.4^{-46} to 3.2^{-43}. When the translation of CALR was searched against protein databases such as SwissProt and PIR, no exact matches were found. Fig 2 shows the comparison of the human calr sequence with that of the dog C5a receptor, CFOMC5AM. (Specification, page 16, line 29 through page 17, line 1.)

(6) THE FINAL REJECTIONS

Claims 12-17 stand rejected under 35 U.S.C. §§ 101 and 112, first paragraph, based on the allegation that the claimed invention lacks patentable utility. The rejection alleges in particular that the invention has "no apparent or disclosed specific and substantial credible utility." (Final Office Action, page 3.)

(7) ISSUES

1. Whether Claims 12-17 directed to a polynucleotide sequence encoding a C5a-like receptor meet the utility requirement of 35 U.S.C. §101.

2. Whether one of ordinary skill in the art would know how to use the claimed polynucleotide, e.g., in toxicology testing, drug development, and the diagnosis of disease, so as to satisfy the enablement requirement of 35 U.S.C. §112, first paragraph.

(8) GROUPING OF THE CLAIMS

As to Issue 1

All of the claims on appeal are grouped together.

As to Issue 2

All of the claims on appeal are grouped together.

(9) APPELLANTS' ARGUMENTS

The rejection of Claims 12-17 is improper, as the inventions of those claims have a patentable utility as set forth in the instant specification, and/or a utility well known to one of ordinary skill in the art.

The invention at issue is a polynucleotide corresponding to a gene that is expressed in a human mast cell library established from the peripheral blood of a patient with mast cell leukemia. The novel polynucleotide codes for a polypeptide demonstrated in the patent specification to be a member of the class of C5a-like seven transmembrane receptors, whose biological functions include binding complement and activating the immune function of mast cells. (Specification, e.g., at page 1, line 3 through page 3, line 16; page 6, lines 7-14; page 16, line 29 through page 17, line 1.) As such, the claimed invention has numerous practical, beneficial uses in drug development, and the diagnosis of disease. As a result of the benefits of these uses, the claimed invention already enjoys significant commercial success.

The Patent Examiner contends that the claimed polynucleotide cannot be useful without precise

knowledge of its biological function. But the law never has required knowledge of biological function to prove utility. It is the claimed invention's uses, not its functions, that are the subject of a proper analysis under the utility requirement.

I. The Applicable Legal Standard

To meet the utility requirement of sections 101 and 112 of the Patent Act, the patent applicant need only show that the claimed invention is "practically useful," *Anderson v. Natta*, 480 F.2d 1392, 1397, 178 USPQ 458 (CCPA 1973) and confers a "specific benefit" on the public. *Brenner v. Manson*, 383 U.S. 519, 534-35, 148 USPQ 689 (1966). As discussed in a recent Court of Appeals for the Federal Circuit case, this threshold is not high:

> An invention is "useful" under section 101 if it is capable of providing some identifiable benefit. See *Brenner v. Manson*, 383 U.S. 519, 534 [148 USPQ 689] (1966); *Brooktree Corp. v. Advanced Micro Devices, Inc.*, 977 F.2d 1555, 1571 [24 USPQ2d 1401] (Fed. Cir. 1992) ("to violate Section 101 the claimed device must be totally incapable of achieving a useful result"); *Fuller v. Berger*, 120 F. 274, 275 (7th Cir. 1903) (test for utility is whether invention "is incapable of serving any beneficial end"). Juicy Whip Inc. v. Orange Bang Inc., 51 USPQ2d 1700 (Fed. Cir. 1999).

While an asserted utility must be described with specificity, the patent applicant need not demonstrate utility to a certainty. In *Stiftung v. Renishaw PLC*, 945 F.2d 1173, 1180, 20 USPQ2d 1094 (Fed. Cir. 1991), the United States Court of Appeals for the Federal Circuit explained:

> An invention need not be the best or only way to accomplish a certain result, and it need only be useful to some extent and in certain applications: "[T]he fact that an invention has only limited utility and is only operable in certain applications is not grounds for finding lack of utility." *Envirotech Corp. v. Al George, Inc.*, 730 F.2d 753, 762, 221 USPQ 473, 480 (Fed. Cir. 1984).

The specificity requirement is not, therefore, an onerous one. If the asserted utility is described so that a person of ordinary skill in the art would understand how to use the claimed invention, it is sufficiently specific. *See Standard Oil Co. v. Montedison, S.p.a.*, 212 U.S.P.Q. 327, 343 (3d Cir. 1981). The specificity requirement is met unless the asserted utility amounts to a "nebulous expression" such as "biological activity" or "biological properties" that does not convey meaningful information about the utility of what is being claimed. *Cross* v. *Iizuka*, 753 F.2d 1040, 1048 (Fed. Cir. 1985).

In addition to conferring a specific benefit on the public, the benefit must also be "substantial." *Brenner*, 383 U.S. at 534. A "substantial" utility is a practical, "real-world" utility. *Nelson v. Bowler*, 626 F.2d 853, 856, 206 USPQ 881 (CCPA 1980).

If persons of ordinary skill in the art would understand that there is a "well-established" utility for the claimed invention, the threshold is met automatically and the applicant need not make any showing to demonstrate utility. Manual of Patent Examination Procedure at § 706.03(a). Only if there is no "well-established" utility for the claimed invention must the applicant demonstrate the practical benefits of the invention. *Id.*

Once the patent applicant identifies a specific utility, the claimed invention is presumed to possess it. *In re Cortright*, 165 F.3d 1353, 1357, 49 USPQ2d 1464 (Fed. Cir. 1999); *In re Brana*, 51 F.3d 1560, 1566; 34 USPQ2d 1436 (Fed. Cir. 1995). In that case, the Patent Office bears the burden of demonstrating that a person of ordinary skill in the art would reasonably doubt that the asserted utility could be achieved by the claimed invention. *Id.* To do so, the Patent Office must provide evidence or sound scientific reasoning. *See In re Langer*, 503 F.2d 1380, 1391-92, 183 USPQ 288 (CCPA 1974). If and only if the Patent Office makes such a showing, the burden shifts to the applicant to provide rebuttal evidence that would convince the person of ordinary skill that there is sufficient proof of utility. *Brana*, 51 F.3d at 1566. The applicant need only prove a "substantial likelihood" of utility; certainty is not required. *Brenner*, 383 U.S. at 532.

II. The uses of polynucleotides encoding CALR for diagnosis of conditions or diseases characterized by expression of CALR and for drug discovery are sufficient utilities under 35 U.S.C. §§ 101 and 112, first paragraph

The claimed invention meets all of the necessary requirements for establishing a credible utility under the Patent Law: There are "well-established" uses for the claimed invention known to persons of ordinary skill in the art, and there are specific practical and beneficial uses for the invention disclosed in the patent application's specification. Objective evidence further corroborates the credibility of the asserted utilities.

A. The uses of polynucleotides encoding CALR for disease diagnosis are practical uses that confer "specific benefits" to the public

The claimed invention has specific, substantial, real-world utility by virtue of its use in disease diagnosis through gene expression profiling. There is no dispute that the claimed invention is in fact a useful tool in hybridization analysis used to perform gene expression analysis. That is sufficient to establish utility for the claimed polynucleotide.

Nowhere does the Patent Examiner address the fact that, as described on page 6, line 15 through page 7, line 1, and page 9, lines 13-22 of the Specification, the claimed polynucleotides can be used as highly specific hybridization probes in, for example, northerns – probes that without question can be used to measure both the existence and amount of complementary RNA sequences known to be the expression products of the claimed polynucleotides. The claimed invention is not, in that regard, some random sequence whose value as a probe is speculative or would require further research to determine.

Given the fact that the claimed polynucleotide is known to be expressed, its utility as a measuring and analyzing instrument for expression levels is as indisputable as a scale's utility for measuring weight. This use as a measuring tool, regardless of how the expression level data ultimately would be used by a person of ordinary skill in the art, by itself demonstrates that the claimed invention provides an identifiable, real-world benefit that meets the utility requirement. *Raytheon* v. *Roper*, 724 F.2d 951, (Fed. Cir. 1983) (claimed invention need only meet one of its stated objectives to be useful); *In re Cortwright*, 165 F.3d 1353, 1359 (Fed. Cir. 1999) (how the invention works is irrelevant to utility); MPEP § 2107 ("Many research tools such as gas chromatographs, screening assays, and nucleotide sequencing techniques have a clear, specific, and unquestionable utility (e.g., <u>they are useful in analyzing compounds</u>)" (emphasis added)).

Though Appellants need not so prove to demonstrate utility, there can be no reasonable dispute that persons of ordinary skill in the art have numerous uses for information about relative gene expression including, for example, understanding the effects of a potential drug for treating mast cell-associated immune conditions caused by abnormal signal transduction due to anaphylactic or hypersensitive responses, systemic and local infections, traumatic and other tissue damage, hereditary or environmental diseases associated with hypertension, carcinomas, and other physiologic or

pathologic problems. In other words, the person of ordinary skill in the art can derive more information about a potential mast cell-associated immune condition drug candidate or potential toxin with the claimed invention than without it.

B. The use of nucleic acids coding for proteins expressed by humans as tools for drug discovery and the diagnosis of disease is now "well-established"

The technologies made possible by expression profiling and the DNA tools upon which they rely are now well-established. The technical literature recognizes not only the prevalence of these technologies, but also their unprecedented advantages in drug development, testing and safety assessment.

Perret J.J. et al. (1992; Biochem J. 288:911-17; IDS Reference No. 6, incorporated by reference into the instant application; Reference No. 1) state that "[u]ltimately, the availability of the cloned receptors should help the design of pharmacologically active (non-peptide) inhibitors that could be used in syndromes were [*sic*: where] inappropriate complement activation occurs." (Perret, page 917.) The Specification discusses using the polynucleotides "in production of chimeric molecules for selecting agonists, inhibitors or antagonists for design of domain-specific therapeutic molecules." (Specification, page 6, lines 27-29.) In addition the Specification describes the use of polypeptides encoded by the claimed polynucleotides in drug screening, for example, page 23, line 12 through page 24, line 14.

Because the Patent Examiner failed to address or consider the "well-established" utilities for the claimed invention in drug development, and the diagnosis of disease, the Examiner's rejections should be overturned regardless of their merit.

C. The similarity of the polypeptide encoded by the claimed invention to another polypeptide of undisputed utility, as well as the expression of the CALR polypeptide in human mast cells, demonstrates utility

The Examiner alleged that "[t]he instant claims are drawn to a protein of as yet undetermined function or biological significance. There is absolutely no evidence of record or any line of reasoning

that would support a conclusion the [*sic*: that] a protein of the instant invention is associated in any way with the plurality of causally unrelated disorders that are listed on page 6 of the instant specification." (Office Action mailed July 25, 2001, page 3.)

Appellants submit that there is adequate evidence in the Specification, along with what is well known in the art, to provide a "line of reasoning" to support the asserted utility for the claimed polynucleotide. This evidence is provided by not only sequence identity between CALR and canine C5a anaphylatoxin receptor but also the expression of CALR in human mast cells.

The utility of the claimed polynucleotide can be imputed based on the relationship between the polypeptide it encodes, CALR, and another polypeptide of unquestioned utility, canine C5a anaphylatoxin receptor. The two polypeptides have sufficient similarities in their sequences that a person of ordinary skill in the art would recognize more than a reasonable probability that the polypeptide encoded for by the claimed invention has utility similar to canine C5a anaphylatoxin receptor. Appellants need not show any more to demonstrate utility. *In re Brana*, 51 F.3d at 1567.

It is undisputed, and readily apparent from the patent application, that the polypeptide encoded for by the claimed polynucleotide shares 46% sequence identity over 152 amino acid residues (L14 through T165 of CALR) with canine C5a anaphylatoxin receptor. This is more than enough homology to demonstrate a reasonable probability that the utility of canine C5a anaphylatoxin receptor can be imputed to the claimed invention (through the polypeptide it encodes). It is well-known that the probability that two unrelated polypeptides share more than 40% sequence homology over 70 amino acid residues is exceedingly small. Brenner et al., Proc. Natl. Acad. Sci. 95:6073-78 (1998) (Reference No. 2). Given homology in excess of 40% over many more than 70 amino acid residues, the probability that the polypeptide encoded for by the claimed polynucleotide is related to canine C5a anaphylatoxin receptor is, accordingly, very high.

The Examiner must accept the Appellants' demonstration that the homology between the polypeptide encoded for by the claimed invention and canine C5a anaphylatoxin receptor demonstrates utility by a reasonable probability unless the Examiner can demonstrate through evidence or sound scientific reasoning that a person of ordinary skill in the art would doubt utility. See *In re Langer*, 503 F.2d 1380, 1391-92, 183 USPQ 288 (CCPA 1974). The Examiner has not provided sufficient

evidence or sound scientific reasoning to the contrary.

Furthermore, confirmation of Appellants' identification of CALR as a human complement receptor is provided in Ames, R.S. et al. (1996; J. Biol. Chem., 271:20231-20334; "Molecular Cloning and Characterization of the Human Anaphylatoxin C3a Receptor"; Reference No. 3), in which the authors describe a human anaphylatoxin C3a receptor that has 98% sequence similarity to CALR.

As discussed *supra*, Perret et al. (Reference No. 1) describe how the availability of the cloned receptors of this family are useful in drug screening.

In addition the Specification describes how polynucleotides encoding CALR are expressed in human mast cells as well as the importance of mast cells in immune response. The Specification teaches that human mast cells have "an important role in promoting various immune responses and nonspecific inflammatory reactions" and "degranulate and discharge granule contents extracellularly," and further that:

> Mast cell granule contents include histamine, heparin, elastase, cathepsin G, eosinophil chemotactic factors, platelet activating factor, and slow-reacting substance of anaphylaxis. When complement cleavage products 3a, 4a, and 5a bind to their respective receptors on the surface of mast cells and basophils, they are capable of triggering the release of histamine and the other factors without the involvement of IgE. Some of the factors listed above are synthesized by mast cells during the course of hypersensitivity reactions and mediate vaso- and broncho-constriction leading to asthma. These and other mediators released following degranulation are responsible both for allergy symptoms and for immunity against some parasites. (Specification, pages 2-3.)

The human mast cell line in which the claimed polynucleotide is expressed "was established from the peripheral blood of a Mayo Clinic patient with mast cell leukemia." (Specification, page 3.) The canine C5a receptor is "present on neutrophils, macrophages, and mast cells." (Specification, page 1, lines 9-10.)

This disclosure provides adequate support for a "line of reasoning" linking the diseases listed on page 6 of the Specification with the claimed polynucleotide. One of skill in the art would reasonably believe that a receptor, expressed in human mast cells and highly similar to a canine C5a receptor, has utility at least in diagnosis and treatment of mast cell-associated immune conditions.

Therefore, for at least the above reasons, the Specification provides adequate support for the

asserted utility of the claimed polynucleotide.

D. Objective evidence corroborates the utilities of the claimed invention

There is, in fact, no restriction on the kinds of evidence a Patent Examiner may consider in determining whether a "real-world" utility exists. Indeed, "real-world" evidence, such as evidence showing actual use or commercial success of the invention, can demonstrate conclusive proof of utility. *Raytheon v. Roper*, 220 USPQ2d 592 (Fed. Cir. 1983); *Nestle v. Eugene*, 55 F.2d 854, 856, 12 USPQ 335 (6th Cir. 1932). Indeed, proof that the invention is made, used or sold by any person or entity other than the patentee is conclusive proof of utility. *United States Steel Corp.* v. *Phillips Petroleum Co.*, 865 F.2d 1247, 1252, 9 USPQ2d 1461 (Fed. Cir. 1989).

Over the past several years, a vibrant market has developed for databases containing all expressed genes (along with the polypeptide translations of those genes), in particular genes having medical and pharmaceutical significance such as the instant sequence. (Note that the value in these databases is enhanced by their completeness, but each sequence in them is independently valuable.) The databases sold by Appellants' assignee, Incyte, include exactly the kinds of information made possible by the claimed invention, such as tissue and disease associations. Incyte sells its database containing the claimed sequence and millions of other sequences throughout the scientific community, including to pharmaceutical companies who use the information to develop new pharmaceuticals.

Both Incyte's customers and the scientific community have acknowledged that Incyte's databases have proven to be valuable in, for example, the identification and development of drug candidates. As Incyte adds information to its databases, including the information that can be generated only as a result of Incyte's discovery of the claimed polynucleotide and its use of that polynucleotide on cDNA microarrays, the databases become even more powerful tools. Thus the claimed invention adds more than incremental benefit to the drug discovery and development process.

III. The Patent Examiner's Rejections Are Without Merit

Rather than responding to the evidence demonstrating utility, the Examiner attempts to dismiss it altogether by arguing that the disclosed and well-established utilities for the claimed polynucleotide are

not "specific and substantial credible" utilities. (Final Office Action, page 3.) The Examiner is incorrect both as a matter of law and as a matter of fact.

A. The Precise Biological Role Or Function Of An Expressed Polynucleotide Is Not Required To Demonstrate Utility

The Patent Examiner's primary rejection of the claimed invention is based on the ground that, without information as to the precise "biological role" of the claimed invention, the claimed invention's utility is not sufficiently specific. According to the Examiner, it is not enough that a person of ordinary skill in the art could use and, in fact, would want to use the claimed invention to monitor the expression of genes for such applications as the evaluation of a drug's efficacy and toxicity. The Examiner would require, in addition, that the applicant provide a specific and substantial interpretation of the results generated in any given expression analysis.

It may be that specific and substantial interpretations and detailed information on biological function are necessary to satisfy the requirements for publication in some technical journals, but they are not necessary to satisfy the requirements for obtaining a United States patent. The relevant question is not, as the Examiner would have it, whether it is known how or why the invention works, *In re Cortwright*, 165 F.3d 1353, 1359 (Fed. Cir. 1999), but rather whether the invention provides an "identifiable benefit" in presently available form. *Juicy Whip Inc.* v. *Orange Bang Inc.*, 185 F.3d 1364, 1366 (Fed. Cir. 1999). If the benefit exists, and there is a substantial likelihood the invention provides the benefit, it is useful. There can be no doubt, that the present invention meets this test.

The threshold for determining whether an invention produces an identifiable benefit is low. *Juicy Whip*, 185 F.3d at 1366. Only those utilities that are so nebulous that a person of ordinary skill in the art would not know how to achieve an identifiable benefit and, at least according to the PTO guidelines, so-called "throwaway" utilities that are not directed to a person of ordinary skill in the art at all, do not meet the statutory requirement of utility. Utility Examination Guidelines, 66 Fed. Reg. 1092 (Jan. 5, 2001).

Knowledge of the biological function or role of a biological molecule has never been required to show real-world benefit. In its most recent explanation of its own utility guidelines, the PTO

acknowledged so much (66 F.R. at 1095):

> [T]he utility of a claimed DNA does not necessarily depend on the function of the encoded gene product. A claimed DNA may have specific and substantial utility because, *e.g.*, it hybridizes near a disease-associated gene or it has gene-regulating activity.

By implicitly requiring knowledge of biological function for any claimed nucleic acid, the Examiner has, contrary to law, elevated what is at most an evidentiary factor into an absolute requirement of utility. Rather than looking to the biological role or function of the claimed invention, the Examiner should have looked first to the benefits it is alleged to provide.

B. Membership in a Class of Useful Products Can Be Proof of Utility

Despite the uncontradicted evidence that the claimed polynucleotide encodes a polypeptide in the C5a-like seven transmembrane receptor family, the Examiner refused to impute the utility of the members of the C5a-like seven transmembrane receptor family to CALR. In the Office Action, the Patent Examiner takes the position that, unless Appellants can identify which particular biological function within the class of C5a-like seven transmembrane receptors is possessed by CALR, utility cannot be imputed. To demonstrate utility by membership in the class of C5a-like seven transmembrane receptors, the Examiner would require that all C5a-like seven transmembrane receptors possess a "common" utility.

There is no such requirement in the law. In order to demonstrate utility by membership in a class, the law requires only that the class not contain a substantial number of useless members. So long as the class does not contain a substantial number of useless members, there is sufficient likelihood that the claimed invention will have utility, and a rejection under 35 U.S.C. § 101 is improper. That is true regardless of how the claimed invention ultimately is used and whether or not the members of the class possess one utility or many. *See Brenner* v. *Manson*, 383 U.S. 519, 532 (1966); *Application of Kirk*, 376 F.2d 936, 943 (CCPA 1967).

Membership in a "general" class is insufficient to demonstrate utility only if the class contains a sufficient number of useless members such that a person of ordinary skill in the art could not impute utility by a substantial likelihood. There would be, in that case, a substantial likelihood that the claimed

invention is one of the useless members of the class. In the few cases in which class membership did not prove utility by substantial likelihood, the classes did in fact include predominately useless members. *E.g.*, *Brenner* (man-made steroids); *Kirk* (same); *Natta* (man-made polyethylene polymers).

The Examiner addresses CALR as if the general class in which it is included is not the C5a-like seven transmembrane receptor family, but rather all polynucleotides or all polypeptides, including the vast majority of useless theoretical molecules not occurring in nature, and thus not pre-selected by nature to be useful. While these "general classes" may contain a substantial number of useless members, the C5a-like seven transmembrane receptor family does not. The C5a-like seven transmembrane receptors family is sufficiently specific to rule out any reasonable possibility that CALR would not also be useful like the other members of the family.

Because the Examiner has not presented any evidence that the C5a-like class of seven transmembrane receptors has any, let alone a substantial number, of useless members, the Examiner must conclude that there is a "substantial likelihood" that the CALR encoded by the claimed polynucleotide is useful. It follows that the claimed polynucleotide also is useful.

Even if the Examiner's "common utility" criterion were correct – and it is not – the C5a-like seven transmembrane receptor family would meet it. It is undisputed that known members of the C5a-like seven transmembrane receptor family are seven transmembrane receptors that bind complement and activate the immune function of mast cells. A person of ordinary skill in the art need not know any more about how the claimed invention binds complement and activates the immune function of mast cells to use it, and the Examiner presents no evidence to the contrary. Instead, the Examiner makes the conclusory observation that a person of ordinary skill in the art would need to know whether, for example, any given C5a-like seven transmembrane receptors binds complement and activates the immune function of mast cells. The Examiner then goes on to assume that the only use for CALR absent knowledge as to how the C5a-like seven transmembrane receptor actually works is further study of CALR itself.

Not so. As demonstrated by Appellants, knowledge that CALR is a C5a-like seven transmembrane receptor is more than sufficient to make it useful for the diagnosis and treatment of mast cell-associated immune conditions. Indeed, CALR has been shown to be expressed in mast cells. The

Examiner must accept these facts to be true unless the Examiner can provide evidence or sound scientific reasoning to the contrary. But the Examiner has not done so.

C. Because the uses of polynucleotides encoding CALR in drug discovery, and disease diagnosis are practical uses beyond mere study of the invention itself, the claimed invention has substantial utility.

The PTO's rejection is tantamount to a rejection based on the polynucleotide being only a research tool and that the use of an invention as a tool for research is not a "substantial" use. Because the PTO's rejection assumes a substantial overstatement of the law, and is incorrect in fact, it must be overturned.

There is no authority for the proposition that use as a tool for research is not a substantial utility. Indeed, the Patent Office has recognized that just because an invention is used in a research setting does not mean that it lacks utility (MPEP § 2107):

> Many research tools such as gas chromatographs, screening assays, and nucleotide sequencing techniques have a clear, specific and unquestionable utility (e.g., they are useful in analyzing compounds). An assessment that focuses on whether an invention is useful only in a research setting thus does not address whether the specific invention is in fact "useful" in a patent sense. Instead, Office personnel must distinguish between inventions that have a specifically identified utility and inventions whose specific utility requires further research to identify or reasonably confirm.

The Patent Office's actual practice has been, at least until the present, consistent with that approach. It has routinely issued patents for inventions whose only use is to facilitate research, such as DNA ligases. These are acknowledged by the PTO's Training Materials themselves to be useful, as well as DNA sequences used, for example, as markers.

Only a limited subset of research uses are not "substantial" utilities: those in which the only known use for the claimed invention is to be an object of further study, thus merely inviting further research. This follows from *Brenner*, in which the U.S. Supreme Court held that a process for making a compound does not confer a substantial benefit where the only known use of the compound was to be the object of further research to determine its use. *Id.* at 535. Similarly, in *Kirk*, the Court held that a compound would not confer substantial benefit on the public merely because it might be used to synthesize some other, unknown compound that would confer substantial benefit. *Kirk*, 376 F.2d at

940, 945 ("What appellants are really saying to those in the art is take these steroids, experiment, and find what use they do have as medicines."). Nowhere do those cases state or imply, however, that a material cannot be patentable if it has some other beneficial use in research.

As used in drug discovery and disease diagnosis, the claimed invention has a beneficial use in research other than studying the claimed invention or its protein products. It is a tool, rather than an object, of research. The data generated in gene expression monitoring using the claimed invention as a tool is not used merely to study the claimed polynucleotide itself, but rather to study properties of tissues, cells, and potential drug candidates and toxins. Without the claimed invention, the information regarding the properties of tissues, cells, drug candidates and toxins is less complete.

The claimed invention has numerous additional uses as a research tool, each of which alone is a "substantial utility." These include uses in chromosomal mapping (Specification, page 9, line 23 through page 10, line 6.)

D. The Patent Examiner Failed to Demonstrate That a Person of Ordinary Skill in the Art Would Reasonably Doubt the Utility of the Claimed Invention

1. Drug screening is a specific, substantial and credible utility

The Examiner argues that "[e]ven if the expression of Applicant's individual protein is affected by a test compound in an array for drug screening, the specification does not disclose any specific and substantial interpretation for the result, and none is known in the art. Given this consideration, the individually claimed antibody[1] has no 'well-established' use." (Final Office Action, page 5.)

Contrary to the Examiner's allegation, there is indeed a "specific and substantial" interpretation for the results of drug screening and toxicology testing using the claimed polynucleotide. Monitoring the expression of the claimed polynucleotide is a method of testing the toxicology of drug candidates during the drug development process. If the expression of a particular polynucleotide is affected in any way by exposure to a test compound, and if that particular polynucleotide (or its encoded polypeptide) is

---

[1]The Examiner in the quoted sentence refers to "the individually claimed antibody." Appellants note that the claims on appeal are directed to a polynucleotide and assume that the Examiner's reference to the "antibody" was made inadvertently.

not the specific target of the test compound (e.g., if the test compound is a drug candidate), then the change in expression is an indication that the test compound has undesirable toxic side effects that may limit its usefulness as a specific drug. Learning this from an array in a gene expression monitoring experiment early in the drug development process costs less than learning this, for example, during Phase III clinical trials. It is important to note that such an indication of possible toxicity is specific not only for each compound tested, but also for each and every individual polynucleotide whose expression is being monitored.

However, the Examiner continues to view the utility of the claimed polynucleotide in toxicology testing and drug screening as requiring knowledge of either the biological function or disease association of the polynucleotide. The Examiner views toxicology testing as a process to measure the toxicity of a drug candidate only when that drug candidate is specifically targeted to the claimed polynucleotide or its encoded polypeptide, alleging that "Applicant has failed to identify the consequences of identifying a compound which is toxic to a polypeptide encoded by the claimed polynucleotide." (Final Office Action, pages 3-4.). The Examiner has refused to consider that the claimed polynucleotide is useful for measuring the toxicity of drug candidates which are targeted not to the claimed polynucleotide or its encoded polypeptide, but to other polynucleotides or polypeptides. This utility of the claimed polynucleotide does not require any knowledge of the biological function or disease association of the claimed polynucleotide or its encoded polypeptide, and is a specific, substantial and credible utility. The Examiner provides neither evidence nor sound scientific reasoning, only unsupported personal opinion, to support the allegation that knowledge of "biological significance" or "disease association" is required for toxicology testing and drug screening.

2. Irrelevance of disease association or differential expression to utility in toxicology testing

The Examiner asserts that the specification does not disclose an association of the claimed polypeptides with "any disease or disorder," and therefore that "the artisan is required to perform substantial further experimentation on the claimed material itself in order to determine to what 'practical use' any expression information regarding this polynucleotide could be put." (Final Office Action, pages 4 and 5.)

These are irrelevant. Appellants need not demonstrate whether the claimed polynucleotide is associated with disease. Appellants need only demonstrate that the claimed polynucleotide is useful.

The claimed polynucleotide can be used for toxicology testing in drug discovery without any knowledge of disease association. Monitoring the expression of the claimed polynucleotide gives important information on the potential toxicity of a drug candidate that is specifically targeted to any other polynucleotide or its encoded polypeptide, regardless of the disease association of the claimed polynucleotide. The claimed polynucleotide is useful for measuring the toxicity of drug candidates specifically targeted to other polynucleotides or their encoded polypeptides, regardless of any possible utility for measuring the properties of the claimed polynucleotide.

3. Discussion of toxicology testing in the Specification

The Examiner alleges that "toxicology testing and drug discovery in the specification as originally filed" and that "the particulars of toxicology testing with SEQ ID NO:2 are not disclosed in the instant specification." (Final Office Action, page 3.) Well-established utilities, such as toxicology testing, need not be explicitly disclosed in a patent application. Furthermore, the Examiner 's position amounts to nothing more than the Examiner's disagreement with Appellants' assertions about the knowledge of a person of ordinary skill. The Examiner must accept Appellants' assertions to be true. The Final Office Action fails to address the disclosure in the instant specification on gene and protein expression monitoring applications, as discussed below.

Support for the utility of the claimed sequences in toxicology testing, as well as for utility in drug screening, may be found in the specification. For example,

> Because CALR is specifically expressed in cells active in immunity, the nucleic acid (calr), polypeptide (CALR) and antibodies to CALR are useful in investigations of and interventions in the normal and abnormal physiologic and pathologic processes which comprise the mast cell's role in immunity. Therefore, an assay for upregulated expression of CALR can accelerate diagnosis and proper treatment of conditions caused by abnormal signal transduction due to anaphylactic or hypersensitive responses, systemic and local infections, traumatic and other tissue damage, hereditary or environmental diseases associated with hypertension, carcinomas, and other physiologic or pathologic problems.
>
> The nucleotide sequence encoding CALR (or its complement) has numerous other applications in techniques known to those skilled in the art of molecular biology.

These techniques include use as hybridization probes for Southerns or northerns, use as oligomers for PCR, use for chromosomal and gene mapping, use in the recombinant production of CALR, use in generation of anti-sense DNA or RNA, their chemical analogs and the like, and use in production of chimeric molecules for selecting agonists, inhibitors or antagonists for design of domain-specific therapeutic molecules. (Coleman '355 application, page 6, lines 15-29.)

The Coleman '355 application further teaches that:

The nucleotide sequence can be used to develop an assay to detect activation, inflammation, or disease associated with abnormal levels of CALR expression. The nucleotide sequence can be labeled by methods known in the art and added to a fluid or tissue sample from a patient. After an incubation period sufficient to effect hybridization, the sample is washed with a compatible fluid which contains a visible marker, a dye or other appropriate molecule(s), if the nucleotide has been labeled with an enzyme. After the compatible fluid is rinsed off, the dye is quantitated and compared with a standard. If the amount of dye is significantly elevated (or lowered, as the case may be), the nucleotide sequence has hybridized with the sample, and the assay indicates an abnormal condition such as inflammation or disease. (Coleman '355 application at page 9, lines 13-22.)

4. Utility of all expressed polypeptides in toxicology testing

The Examiner asserts that use as a control for toxicology testing is not specific and substantial, and therefore not well-established, because it "would apply to virtually every member of a general class of materials, such as any collection of proteins or DNAs, but it is only potential with respect to SEQ ID NO:2." (Final Office Action, page 4). The Examiner does not point to any law, however, that says a utility that is shared by a large class is somehow not a utility. If all of the class of polypeptides or polynucleotides can be so used, then they all have utility. The issue is, once again, whether the claimed invention has any utility, not whether other compounds have a similar utility. Nothing in the law says that an invention must have a "unique" utility. Indeed, the whole notion of "well established" utilities presupposes that many different inventions can have the exact same utility. If the Examiner's argument was correct, there could never be a well established utility, because you could always find a generic group with the same utility!

Furthermore, the Examiner is incorrect in stating that "virtually every member of a general class of materials, such as any collection of proteins or DNAs" could be used in toxicology testing. (Final

Office Action, page 4.) The property of the claimed polynucleotide that makes it useful as a control for toxicology testing is its expression in naturally occurring cells. A polynucleotide having a random, non-naturally occurring sequence would most likely not be useful as a control for toxicology testing.

The Examiner further asserts that "the information that is gained from the array is dependent on the pattern derived from the array, and says nothing with regard to each individual member of the array" and that this is, again, a general utility. (Final Office Action, pages 4-5.) Appellants note that while the information derived from an array does depend upon the pattern derived from individual members of the array, an array still cannot be made without individual members. Thus each individual polynucleotide sequence has a utility in creating arrays. Each of these individual polynucleotide sequences has a unique and specific utility in that it records the expression level of a unique gene. This is a substantial, "real world" utility in that one of ordinary skill in the art would know how to use the claimed polynucleotide's sequence in an array, without any further experimentation.

IV. By Requiring the Patent Applicant to Assert a Particular or Unique Utility, the Patent Examination Utility Guidelines and Training Materials Applied by the Patent Examiner Misstate the Law

There is an additional, independent reason to overturn the rejections: to the extent the rejections are based on Revised Interim Utility Examination Guidelines (64 FR 71427, December 21, 1999), the final Utility Examination Guidelines (66 FR 1092, January 5, 2001) and/or the Revised Interim Utility Guidelines Training Materials (USPTO Website www.uspto.gov, March 1, 2000), the Guidelines and Training Materials are themselves inconsistent with the law.

The Training Materials, which direct the Examiners regarding how to apply the Utility Guidelines, address the issue of specificity with reference to two kinds of asserted utilities: "specific" utilities which meet the statutory requirements, and "general" utilities which do not. The Training Materials define a "specific utility" as follows:

> A [specific utility] is *specific* to the subject matter claimed. This contrasts to *general* utility that would be applicable to the broad class of invention. For example, a claim to a polynucleotide whose use is disclosed simply as "gene probe" or "chromosome marker" would not be considered to be specific in the absence of a disclosure of a specific DNA target. Similarly, a general statement of diagnostic utility, such as

diagnosing an unspecified disease, would ordinarily be insufficient absent a disclosure of what condition can be diagnosed.

The Training Materials distinguish between "specific" and "general" utilities by assessing whether the asserted utility is sufficiently "particular," *i.e.*, unique (Training Materials at p.52) as compared to the "broad class of invention." (In this regard, the Training Materials appear to parallel the view set forth in Stephen G. Kunin, Written Description Guidelines and Utility Guidelines, 82 J.P.T.O.S. 77, 97 (Feb. 2000) ("With regard to the issue of specific utility the question to ask is whether or not a utility set forth in the specification is *particular* to the claimed invention.")).

Such "unique" or "particular" utilities never have been required by the law. To meet the utility requirement, the invention need only be "practically useful," *Natta*, 480 F.2d 1 at 1397, and confer a "specific benefit" on the public. *Brenner*, 383 U.S. at 534. Thus, incredible "throwaway" utilities, such as trying to "patent a transgenic mouse by saying it makes great snake food," do not meet this standard. Karen Hall, Genomic Warfare, The American Lawyer 68 (June 2000) (quoting John Doll, Chief of the Biotech Section of USPTO).

This does not preclude, however, a general utility, contrary to the statement in the Training Materials where "specific utility" is defined (page 5). Practical real-world uses are not limited to uses that are unique to an invention. The law requires that the practical utility be "definite," not particular. *Montedison,* 664 F.2d at 375. Appellant is not aware of any court that has rejected an assertion of utility on the grounds that it is not "particular" or "unique" to the specific invention. Where courts have found utility to be too "general," it has been in those cases in which the asserted utility in the patent disclosure was not a practical use that conferred a specific benefit. That is, a person of ordinary skill in the art would have been left to guess as to how to benefit at all from the invention. In *Kirk*, for example, the CCPA held the assertion that a man-made steroid had "useful biological activity" was insufficient where there was no information in the specification as to how that biological activity could be practically used. *Kirk*, 376 F.2d at 941.

The fact that an invention can have a particular use does not provide a basis for requiring a particular use. *See Brana, supra* (disclosure describing a claimed antitumor compound as being homologous to an antitumor compound having activity against a "particular" type of cancer was determined to satisfy the specificity requirement). "Particularity" is not and never has been the *sine qua*

non of utility; it is, at most, one of many factors to be considered.

As described *supra*, broad classes of inventions can satisfy the utility requirement so long as a person of ordinary skill in the art would understand how to achieve a practical benefit from knowledge of the class. Only classes that encompass a significant portion of nonuseful members would fail to meet the utility requirement. *Supra* § II.B. (*Montedison,* 664 F.2d at 374-75).

The Training Materials fail to distinguish between broad classes that convey information of practical utility and those that do not, lumping all of them into the latter, unpatentable category of "general" utilities. As a result, the Training Materials paint with too broad a brush. Rigorously applied, they would render unpatentable whole categories of inventions that heretofore have been considered to be patentable and that have indisputably benefitted the public, including the claimed invention. *See supra* § II.B. Thus the Training Materials cannot be applied consistently with the law.

V. To the Extent the Rejection of the Claimed Invention under 35 U.S.C. § 112, First Paragraph, Is Based on the Improper Rejection for Lack of Utility under 35 U.S.C. § 101, it Must Be Reversed.

The rejection set forth in the Office Action is based on the assertions discussed above, i.e., that the claimed invention lacks patentable utility. To the extent that the rejection under § 112, first paragraph, is based on the improper allegation of lack of patentable utility under § 101, it fails for the same reasons.

(10) CONCLUSION

Appellants respectfully submit that rejections for lack of utility based, *inter alia*, on an allegation of "lack of specificity," as set forth in the Office Action and as justified in the Revised Interim and final Utility Guidelines and Training Materials, are not supported in the law. Neither are they scientifically correct, nor supported by any evidence or sound scientific reasoning. These rejections are alleged to be founded on facts in court cases such as *Brenner* and *Kirk*, yet those facts are clearly distinguishable from the facts of the instant application, and indeed most if not all nucleotide and protein sequence applications. Nevertheless, the PTO is attempting to mold the facts and holdings of these

prior cases, "like a nose of wax," [2]to target rejections of claims to polypeptides and polynucleotides where biological activity information has not been proven by laboratory experimentation, and they have done so by ignoring perfectly acceptable utilities fully disclosed in the specification as well as well-established utilities known to those of skill in the art. As is disclosed in the specification, and even more clearly, as one of ordinary skill in the art would understand, the claimed invention has well-established, specific, substantial and credible utilities. The rejections are, therefore, improper and should be reversed.

Moreover, to the extent the above rejections were based on the Revised Interim and final Examination Guidelines and Training Materials, those portions of the Guidelines and Training Materials that form the basis for the rejections should be determined to be inconsistent with the law.

Due to the urgency of this matter, including its economic and public health implications, an expedited review of this appeal is earnestly solicited.

If the USPTO determines that any additional fees are due, the Commissioner is hereby authorized to charge Deposit Account No. **09-0108.**

This brief is enclosed in triplicate.

Respectfully submitted,

INCYTE GENOMICS, INC.

Date: October 2, 2002

Susan K. Sather
Reg. No. 44,316
Direct Dial Telephone: (650) 845-4646

3160 Porter Drive
Palo Alto, California 94304
Phone: (650) 855-0555
Fax: (650) 849-8886

[2]"The concept of patentable subject matter under §101 is not 'like a nose of wax which may be turned and twisted in any direction * * *.' *White v. Dunbar*, 119 U.S. 47, 51." *(Parker v. Flook,* 198 USPQ 193 (US SupCt 1978))

APPENDIX - CLAIMS ON APPEAL

12. (As Once Amended) An isolated polynucleotide comprising a polynucleotide sequence encoding the amino acid sequence of SEQ ID NO:2.

13. (As Once Amended) An isolated polynucleotide comprising the polynucleotide sequence of SEQ ID NO:1.

14. (As Once Amended) An isolated polynucleotide fully complementary to a polynucleotide comprising the polynucleotide sequence of SEQ ID NO:1.

15. (As Once Amended) An expression vector comprising the isolated polynucleotide of claim 12.

16. (Reiterated) A host cell comprising the expression vector of claim 15.

17. (As Once Amended) A method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:2, said method comprising the steps of:

(a) culturing the host cell of claim 16 under conditions suitable for expression of the polypeptide, and

(b) recovering said polypeptide from the cell culture.

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 6073–6078, May 1998
Biochemistry

Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships

STEVEN E. BRENNER*†‡, CYRUS CHOTHIA*, AND TIM J. P. HUBBARD§

*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom; and §Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambs CB10 1SA, United Kingdom

Communicated by David R. Davies, National Institute of Diabetes, Bethesda, MD, March 16, 1998 (received for review November 12, 1997)

ABSTRACT **Pairwise sequence comparison methods have been assessed using proteins whose relationships are known reliably from their structures and functions, as described in the SCOP database [Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia C. (1995) *J. Mol. Biol.* 247, 536–540]. The evaluation tested the programs BLAST [Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* 215, 403–410], WU-BLAST2 [Altschul, S. F. & Gish, W. (1996) *Methods Enzymol.* 266, 460–480], FASTA [Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2444–2448], and SSEARCH [Smith, T. F. & Waterman, M. S. (1981) *J. Mol. Biol.* 147, 195–197] and their scoring schemes. The error rate of all algorithms is greatly reduced by using statistical scores to evaluate matches rather than percentage identity or raw scores. The E-value statistical scores of SSEARCH and FASTA are reliable: the number of false positives found in our tests agrees well with the scores reported. However, the P-values reported by BLAST and WU-BLAST2 exaggerate significance by orders of magnitude. SSEARCH, FASTA ktup = 1, and WU-BLAST2 perform best, and they are capable of detecting almost all relationships between proteins whose sequence identities are >30%. For more distantly related proteins, they do much less well; only one-half of the relationships between proteins with 20–30% identity are found. Because many homologs have low sequence similarity, most distant relationships cannot be detected by any pairwise comparison method; however, those which are identified may be used with confidence.**

Sequence database searching plays a role in virtually every branch of molecular biology and is crucial for interpreting the sequences issuing forth from genome projects. Given the method's central role, it is surprising that overall and relative capabilities of different procedures are largely unknown. It is difficult to verify algorithms on sample data because this requires large data sets of proteins whose evolutionary relationships are known unambiguously and independently of the methods being evaluated. However, nearly all known homologs have been identified by sequence analysis (the method to be tested). Also, it is generally very difficult to know, in the absence of structural data, whether two proteins that lack clear sequence similarity are unrelated. This has meant that although previous evaluations have helped improve sequence comparison, they have suffered from insufficient, imperfectly characterized, or artificial test data. Assessment also has been problematic because high quality database sequence searching attempts to have both sensitivity (detection of homologs) and specificity (rejection of unrelated proteins); however, these complementary goals are linked such that increasing one causes the other to be reduced.

Sequence comparison methodologies have evolved rapidly, so no previously published tests has evaluated modern versions of programs commonly used. For example, parameters in BLAST (1) have changed, and WU-BLAST2 (2)—which produces gapped alignments—has become available. The latest version of FASTA (3) previously tested was 1.6, but the current release (version 3.0) provides fundamentally different results in the form of statistical scoring.

The previous reports also have left gaps in our knowledge. For example, there has been no published assessment of thresholds for scoring schemes more sophisticated than percentage identity. Thus, the widely discussed statistical scoring measures have never actually been evaluated on large databases of real proteins. Moreover, the different scoring schemes commonly in use have not been compared.

Beyond these issues, there is a more fundamental question: in an absolute sense, how well does pairwise sequence comparison work? That is, what fraction of homologous proteins can be detected using modern database searching methods?

In this work, we attempt to answer these questions and to overcome both of the fundamental difficulties that have hindered assessment of sequence comparison methodologies. First, we use the set of distant evolutionary relationships in the SCOP: Structural Classification of Proteins database (4), which is derived from structural and functional characteristics (5). The SCOP database provides a uniquely reliable set of homologs, which are known independently of sequence comparison. Second, we use an assessment method that jointly measures both sensitivity and specificity. This method allows straightforward comparison of different sequence searching procedures. Further, it can be used to aid interpretation of real database searches and thus provide optimal and reliable results.

Previous Assessments of Sequence Comparison. Several previous studies have examined the relative performance of different sequence comparison methods. The most encompassing analyses have been by Pearson (6, 7), who compared the three most commonly used programs. Of these, the Smith–Waterman algorithm (8) implemented in SSEARCH (3) is the oldest and slowest but the most rigorous. Modern heuristics have provided BLAST (1) the speed and convenience to make it the most popular program. Intermediate between these two is FASTA (3), which may be run in two modes offering either greater speed (ktup = 2) or greater effectiveness (ktup = 1). Pearson also considered different parameters for each of these programs.

To test the methods, Pearson selected two representative proteins from each of 67 protein superfamilies defined by the PIR database (9). Each was used as a query to search the database, and the matched proteins were marked as being homologous or unrelated according to their membership of PIR

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: EPQ, errors per query.

†Present address: Department of Structural Biology, Stanford University, Fairchild Building D-109, Stanford, CA 94305-5126
‡To whom reprints requests should be addressed. e-mail: brenner@hyper.stanford.edu.

superfamilies. Pearson found that modern matrices and "lnscaling" of raw scores improve results considerably. He also reported that the rigorous Smith–Waterman algorithm worked slightly better than FASTA, which was in turn more effective than BLAST.

Very large scale analyses of matrices have been performed (10), and Henikoff and Henikoff (11) also evaluated the effectiveness of BLAST and FASTA. Their test with BLAST considered the ability to detect homologs above a predetermined score but had no penalty for methods which also reported large numbers of spurious matches. The Henikoffs searched the SWISS-PROT database (12) and used PROSITE (13) to define homologous families. Their results showed that the BLOSUM62 matrix (14) performed markedly better than the extrapolated PAM-series matrices (15), which previously had been popular.

A crucial aspect of any assessment is the data that are used to test the ability of the program to find homologs. But in Pearson's and the Henikoffs' evaluations of sequence comparison, the correct results were effectively unknown. This is because the superfamilies in PIR and PROSITE are principally created by using the same sequence comparison methods which are being evaluated. Interdependency of data and methods creates a "chicken and egg" problem, and means for example, that new methods would be penalized for correctly identifying homologs missed by older programs. For instance, immunoglobulin variable and constant domains are clearly homologous, but PIR places them in different superfamilies. The problem is widespread: each superfamily in PIR 48.00 with a structural homolog is itself homologous to an average of 1.6 other PIR superfamilies (16).

To surmount these sorts of difficulties, Sander and Schneider (17) used protein structures to evaluate sequence comparison. Rather than comparing different sequence comparison algorithms, their work focused on determining a length-dependent threshold of percentage identity, above which all proteins would be of similar structure. A result of this analysis was the HSSP equation; it states that proteins with 25% identity over 80 residues will have similar structures, whereas shorter alignments require higher identity. (Other studies also have used structures (18–20), but these focused on a small number of model proteins and were principally oriented toward evaluating alignment accuracy rather than homology detection.)

A general solution to the problem of scoring comes from statistical measures (i.e., E-values and P-values) based on the extreme value distribution (21). Extreme value scoring was implemented analytically in the BLAST program using the Karlin and Altschul statistics (22, 23) and empirical approaches have been recently added to FASTA and SSEARCH. In addition to being heralded as a reliable means of recognizing significantly similar proteins (24, 25), the mathematical tractability of statistical scores "is a crucial feature of the BLAST algorithm" (1). The validity of this scoring procedure has been tested analytically and empirically (see ref. 2 and references in ref. 24). However, all large empirical tests used random sequences that may lack the subtle structure found within biological sequences (26, 27) and obviously do not contain any real homologs. Thus, although many researchers have suggested that statistical scores be used to rank matches (24, 25, 28), there have been no large rigorous experiments on biological data to determine the degree to which such rankings are superior.

A Database for Testing Homology Detection. Since the discovery that the structures of hemoglobin and myoglobin are very similar though their sequences are not (29), it has been apparent that comparing structures is a more powerful (if less convenient) way to recognize distant evolutionary relationships than comparing sequences. If two proteins show a high degree of similarity in their structural details and function, it is very probable that they have an evolutionary relationship though their sequence similarity may be low.

The recent growth of protein structure information combined with the comprehensive evolutionary classification in the SCOP database (4, 5) have allowed us to overcome previous limitations. With these data, we can evaluate the performance of sequence comparison methods on real protein sequences whose relationships are known confidently. The SCOP database uses structural information to recognize distant homologs, the large majority of which can be determined unambiguously. These superfamilies, such as the globins or the immunoglobulins, would be recognized as related by the vast majority of the biological community despite the lack of high sequence similarity.

From SCOP, we extracted the sequences of domains of proteins in the Protein Data Bank (PDB) (30) and created two databases. One (PDB90D-B) has domains, which were all <90% identical to any other, whereas (PDB40D-B) had those <40% identical. The databases were created by first sorting all protein domains in SCOP by their quality and making a list. The highest quality domain was selected for inclusion in the database and removed from the list. Also removed from the list (and discarded) were all other domains above the threshold level of identity to the selected domain. This process was repeated until the list was empty. The PDB40D-B database contains 1,323 domains, which have 9,044 ordered pairs of distant relationships, or ~0.5% of the total 1,749,006 ordered pairs. In PDB90D-B, the 2,079 domains have 53,988 relationships, representing 1.2% of all pairs. Low complexity regions of sequence can achieve spurious high scores, so these were masked in both databases by processing with the SEG program (27) using recommended parameters: 12 1.8 2.0. The databases used in this paper are available from http://sss.stanford.edu/sss/, and databases derived from the current version of SCOP may be found at http://scop.mrc-lmb.cam.ac.uk/scop/.

Analyses from both databases were generally consistent, but PDB40D-B focuses on distantly related proteins and reduces the heavy overrepresentation in the PDB of a small number of families (31, 32), whereas PDB90D-B (with more sequences) improves evaluations of statistics. Except where noted otherwise, the distant homolog results here are from PDB40D-B. Although the precise numbers reported here are specific to the structural domain databases used, we expect the trends to be general.

Assessment Data and Procedure. Our assessment of sequence comparison may be divided into four different major categories of tests. First, using just a single sequence comparison algorithm at a time, we evaluated the effectiveness of different scoring schemes. Second, we assessed the reliability of scoring procedures, including an evaluation of the validity of statistical scoring. Third, we compared sequence comparison algorithms (using the optimal scoring scheme) to determine their relative performance. Fourth, we examined the distribution of homologs and considered the power of pairwise sequence comparison to recognize them. All of the analyses used the databases of structurally identified homologs and a new assessment criterion.

The analyses tested BLAST (1), version 1.4.9MP, and WU-BLAST2 (2), version 2.0a13MP. Also assessed was the FASTA package, version 3.0t76 (3), which provided FASTA and the SSEARCH implementation of Smith–Waterman (8). For SSEARCH and FASTA, we used BLOSUM45 with gap penalties −12/−1 (7, 16). The default parameters and matrix (BLOSUM62) were used for BLAST and WU-BLAST2.

The "Coverage Vs. Error" Plot. To test a particular protocol (comprising a program and scoring scheme), each sequence from the database was used as a query to search the database. This yielded ordered pairs of query and target sequences with associated scores, which were sorted, on the basis of their scores, from best to worst. The ideal method would have

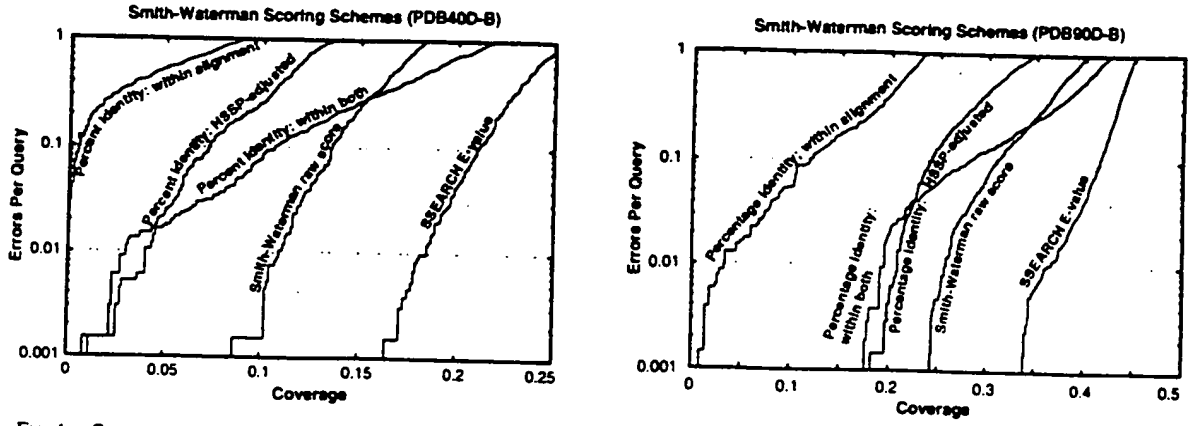


FIG. 1. Coverage vs. error plots of different scoring schemes for SSEARCH Smith–Waterman. (*A*) Analysis of PDB40D-B database. (*B*) Analysis of PDB90D-B database. All of the proteins in the database were compared with each other using the SSEARCH program. The results of this single set of comparisons were considered using five different scoring schemes and assessed. The graphs show the coverage and errors per query (EPQ) for statistical scores, raw scores, and three measures using percentage identity. In the coverage vs. error plot, the *x* axis indicates the fraction of all homologs in the database (known from structure) which have been detected. Precisely, it is the number of detected pairs of proteins with the same fold divided by the total number of pairs from a common superfamily. PDB40D-B contains a total of 9,044 homologs, so a score of 10% indicates identification of 904 relationships. The *y* axis reports the number of EPQ. Because there are 1,323 queries made in the PDB40D-B all-vs.-all comparison, 13 errors corresponds to 0.01, or 1% EPQ. The *y* axis is presented on a log scale to show results over the widely varying degrees of accuracy which may be desired. The scores that correspond to the levels of EPQ and coverage are shown in Fig. 4 and Table 1. The graph demonstrates the trade-off between sensitivity and selectivity. As more homologs are found (moving to the right), more errors are made (moving up). The ideal method would be in the lower right corner of the graph, which corresponds to identifying many evolutionary relationships without selecting unrelated proteins. Three measures of percentage identity are plotted. Percentage identity within alignment is the degree of identity within the aligned region of the proteins, without consideration of the alignment length. Percentage identity within both is the number of identical residues in the aligned region as a percentage of the average length of the query and target proteins. The HSSP equation (17) is $H = 290.15l^{-0.562}$ where l is length for $10 < l < 80$; $H > 100$ for $l < 10$; $H = 24.7$ for $l > 80$. The percentage identity HSSP-adjusted score is the percent identity within the alignment minus H. Smith–Waterman raw scores and E-values were taken directly from the sequence comparison program.

perfect separation, with all of the homologs at the top of the list and unrelated proteins below. In practice, perfect separation is impossible to achieve so instead one is interested in drawing a threshold above which there are the largest number of related pairs of sequences consistent with an acceptable error rate.

Our procedure involved measuring the coverage and error for every threshold. Coverage was defined as the fraction of structurally determined homologs that have scores above the selected threshold; this reflects the sensitivity of a method. Errors per query (EPQ), an indicator of selectivity, is the number of nonhomologous pairs above the threshold divided by the number of queries. Graphs of these data, called coverage vs. error plots, were devised to understand how protocols compare at different levels of accuracy. These graphs share effectively all of the beneficial features of Reciever Operating Characteristic (ROC) plots (33, 34) but better represent the high degrees of accuracy required in sequence comparison and the huge background of nonhomologs.

This assessment procedure is directly relevant to practical sequence database searching, for it provides precisely the information necessary to perform a reliable sequence database search. The EPQ measure places a premium on score consistency; that is, it requires scores to be comparable for different queries. Consistency is an aspect which has been largely

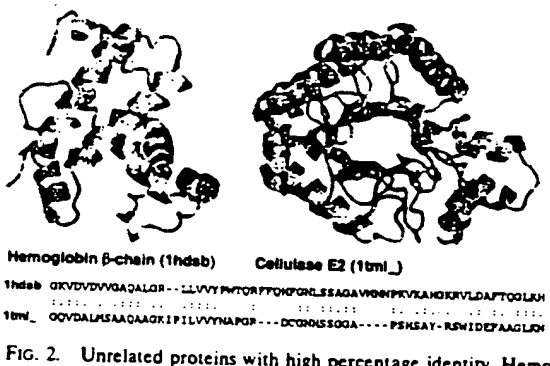


FIG. 2. Unrelated proteins with high percentage identity. Hemoglobin β-chain (PDB code 1hds chain b, ref. 38, *Left*) and cellulase E2 (PDB code 1tml, ref. 39, *Right*) have 39% identity over 64 residues, a level which is often believed to be indicative of homology. Despite this high degree of identity, their structures strongly suggest that these proteins are not related. Appropriately, neither the raw alignment score of 85 nor the E-value of 1.3 is significant. Proteins rendered by RASMOL (40).

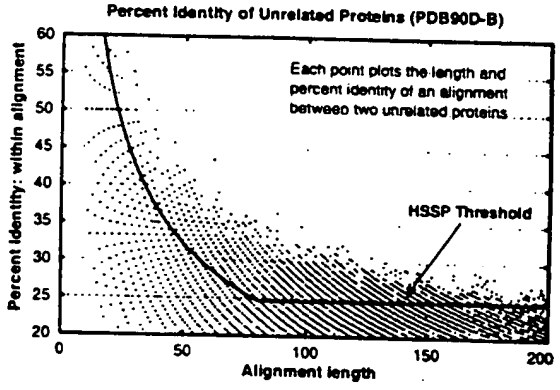


FIG. 3. Length and percentage identity of alignments of unrelated proteins in PDB90D-B: Each pair of nonhomologous proteins found with SSEARCH is plotted as a point whose position indicates the length and the percentage identity within the alignment. Because alignment length and percentage identity are quantized, many pairs of proteins may have exactly the same alignment length and percentage identity. The line shows the HSSP threshold (though it is intended to be applied with a different matrix and parameters).

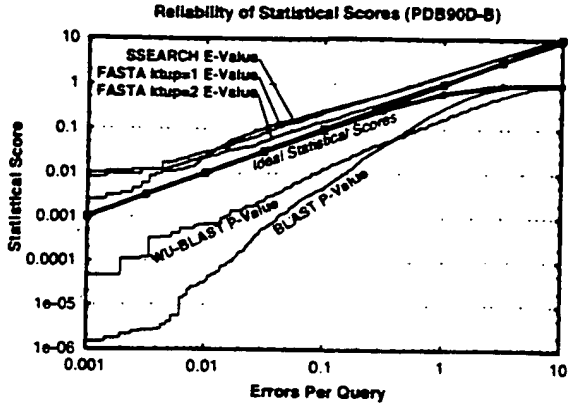


FIG. 4. Reliability of statistical scores in PDB90D-B: Each line shows the relationship between reported statistical score and actual error rate for a different program. E-values are reported for SSEARCH and FASTA, whereas P-values are shown for BLAST and WU-BLAST2. If the scoring were perfect, then the number of errors per query and the E-values would be the same, as indicated by the upper bold line. (P-values should be the same as EPQ for small numbers, and diverges at higher values, as indicated by the lower bold line.) E-values from SSEARCH and FASTA are shown to have good agreement with EPQ but underestimate the significance slightly. BLAST and WU-BLAST2 are overconfident, with the degree of exaggeration dependent upon the score. The results for PDB40D-B were similar to those for PDB90D-B despite the difference in number of homologs detected. This graph could be used to roughly calibrate the reliability of a given statistical score.

ignored in previous tests but is essential for the straightforward or automatic interpretation of sequence comparison results. Further, it provides a clear indication of the confidence that should be ascribed to each match. Indeed, the EPQ measure should approximate the expectation value reported by database searching programs, if the programs' estimates are accurate.

The Performance of Scoring Schemes. All of the programs tested could provide three fundamental types of scores. The first score is the percentage identity, which may be computed in several ways based on either the length of the alignment or the lengths of the sequences. The second is a "raw" or "Smith–Waterman" score, which is the measure optimized by the Smith–Waterman algorithm and is computed by summing the substitution matrix scores for each position in the alignment and subtracting gap penalties. In BLAST, a measure related to this score is scaled into bits. Third is a statistical score based on the extreme value distribution. These results are summarized in Fig. 1.

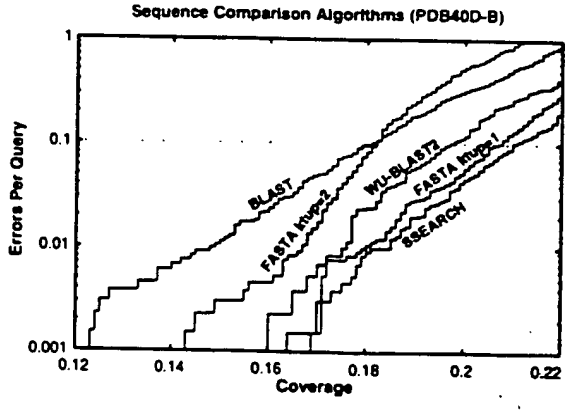
Sequence Identity. Though it has been long established that percentage identity is a poor measure (35), there is a common rule-of-thumb stating that 30% identity signifies homology. Moreover, publications have indicated that 25% identity can be used as a threshold (17, 36). We find that these thresholds, originally derived years ago, are not supported by present results. As databases have grown, so have the possibilities for chance alignments with high identity; thus, the reported cutoffs lead to frequent errors. Fig. 2 shows one of the many pairs of proteins with very different structures that nonetheless have high levels of identity over considerable aligned regions. Despite the high identity, the raw and the statistical scores for such incorrect matches are typically not significant. The principal reasons percentage identity does so poorly seem to be that it ignores information about gaps and about the conservative or radical nature of residue substitutions.

From the PDB90D-B analysis in Fig. 3, we learn that 30% identity is a reliable threshold for this database only for sequence alignments of at least 150 residues. Because one unrelated pair of proteins has 43.5% identity over 62 residues, it is probably necessary for alignments to be at least 70 residues in length before 40% is a reasonable threshold, for a database of this particular size and composition.

At a given reliability, scores based on percentage identity detect just a fraction of the distant homologs found by statistical scoring. If one measures the percentage identity in the aligned regions without consideration of alignment length, then a negligible number of distant homologs are detected. Use of the HSSP equation improves the value of percentage identity, but even this measure can find only 4% of all known homologs at 1% EPQ. In short, percentage identity discards most of the information measured in a sequence comparison.

Raw Scores. Smith–Waterman raw scores perform better than percentage identity (Fig. 1), but ln-scaling (7) provided no notable benefit in our analysis. It is necessary to be very precise when using either raw or bit scores because a 20% change in cutoff score could yield a tenfold difference in EPQ. However, it is difficult to choose appropriate thresholds because the reliability of a bit score depends on the lengths of the proteins matched and the size of the database. Raw score thresholds also are affected by matrix and gap parameters.

Statistical Scores. Statistical scores were introduced partly to overcome the problems that arise from raw scores. This scoring scheme provides the best discrimination between homologous proteins and those which are unrelated. Most



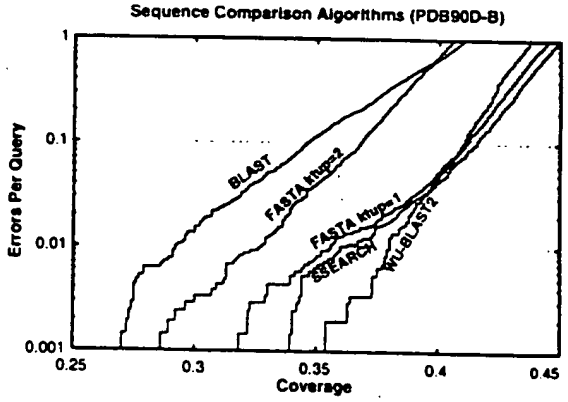


FIG. 5. Coverage vs. error plots of different sequence comparison methods: Five different sequence comparison methods are evaluated, each using statistical scores (E- or P-values). (*A*) PDB40D-B database. In this analysis, the best method is the slow SSEARCH, which finds 18% of relationships at 1% EPQ. FASTA ktup = 1 and WU-BLAST2 are almost as good. (*B*) PDB90D-B database. The quick WU-BLAST2 program provides the best coverage at 1% EPQ on this database, although at higher levels of error it becomes slightly worse than FASTA ktup = 1 and SSEARCH.

likely, its power can be attributed to its incorporation of more information than any other measure; it takes account of the full substitution and gap data (like raw scores) but also has details about the sequence lengths and composition and is scaled appropriately.

We find that statistical scores are not only powerful, but also easy to interpret. SSEARCH and FASTA show close agreement between statistical scores and actual number of errors per query (Fig. 4). The expectation value score gives a good, slightly conservative estimate of the chances of the two sequences being found at random in a given query. Thus, an E-value of 0.01 indicates that roughly one pair of nonhomologs of this similarity should be found in every 100 different queries. Neither raw scores nor percentage identity can be interpreted in this way, and these results validate the suitability of the extreme value distribution for describing the scores from a database search.

The P-values from BLAST also should be directly interpretable but were found to overstate significance by more than two orders of magnitude for 1% EPQ for this database. Nonetheless, these results strongly suggest that the analytic theory is fundamentally appropriate. WU-BLAST2 scores were more reliable than those from BLAST, but also exaggerate expected confidence by more than an order of magnitude at 1% EPQ.

Overall Detection of Homologs and Comparison of Algorithms. The results in Fig. 5*A* and Table 1 show that pairwise sequence comparison is capable of identifying only a small fraction of the homologous pairs of sequences in PDB40D-B. Even SSEARCH with E-values, the best protocol tested, could find only 18% of all relationships at a 1% EPQ. BLAST, which identifies 15%, was the worst performer, whereas FASTA ktup = 1 is nearly as effective as SSEARCH. FASTA ktup = 2 and WU-BLAST2 are intermediate in their ability to detect homologs. Comparison of different algorithms indicates that those capable of identifying more homologs are generally slower. SSEARCH is 25 times slower than BLAST and 6.5 times slower than FASTA ktup = 1. WU-BLAST2 is slightly faster than FASTA ktup = 2, but the latter has more interpretable scores.

In PDB90D-B, where there are many close relationships, the best method can identify only 38% of structurally known homologs (Fig. 5*B*). The method which finds that many relationships is WU-BLAST2. Consequently, we infer that the differences between FASTA kup = 1, SSEARCH, and WU-BLAST2 programs are unlikely to be significant when compared with variation in database composition and scoring reliability.

Fig. 6 helps to explain why most distant homologs cannot be found by sequence comparison: a great many such relationships have no more sequence identity than would be expected by chance. SSEARCH with E-values can recognize >90% of the homologous pairs with 30–40% identity. In this region, there are 30 pairs of homologous proteins that do not have significant E-values, but 26 of these involve sequences with <50 residues. Of sequences having 25–30% identity, 75% are identified by SSEARCH E-values. However, although the number of homologs grows at lower levels of identity, the detection falls off sharply: only 40% of homologs with 20–25% identity

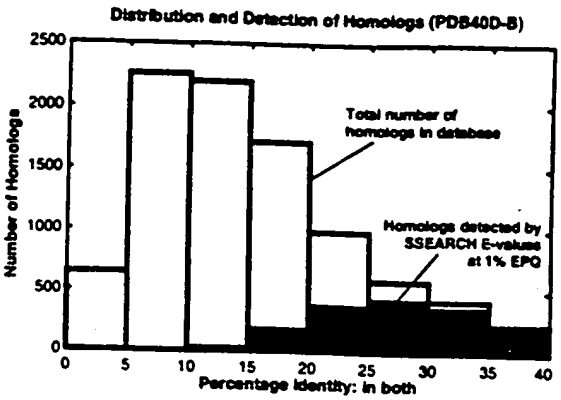


FIG. 6. Distribution and detection of homologs in PDB40D-B. Bars show the distribution of homologous pairs PDB40D-B according to their identity (using the measure of identity in both). Filled regions indicate the number of these pairs found by the best database searching method (SSEARCH with E-values) at 1% EPQ. The PDB40D-B database contains proteins with <40% identity, and as shown on this graph, most structurally identified homologs in the database have diverged extremely far in sequence and have <20% identity. Note that the alignments may be inaccurate, especially at low levels of identity. Filled regions show that SSEARCH can identify most relationships that have 25% or more identity, but its detection wanes sharply below 25%. Consequently, the great sequence divergence of most structurally identified evolutionary relationships effectively defeats the ability of pariwise sequence comparison to detect them.

are detected and only 10% of those with 15–20% can be found. These results show that statistical scores can find related proteins whose identity is remarkably low; however, the power of the method is restricted by the great divergence of many protein sequences.

After completion of this work, a new version of pairwise BLAST was released: BLASTGP (37). It supports gapped alignments, like WU-BLAST2, and dispenses with sum statistics. Our initial tests on BLASTGP using default parameters show that its E-values are reliable and that its overall detection of homologs was substantially better than that of ungapped BLAST, but not quite equal to that of WU-BLAST2.

CONCLUSION

The general consensus amongst experts (see refs. 7, 24, 25, 27 and references therein) suggests that the most effective sequence searches are made by (*i*) using a large current database in which the protein sequences have been complexity masked and (*ii*) using statistical scores to interpret the results. Our experiments fully support this view.

Our results also suggest two further points. First, the E-values reported by FASTA and SSEARCH give fairly accurate estimates of the significance of each match, but the P-values provided by BLAST and WU-BLAST2 underestimate the true

Table 1. Summary of sequence comparison methods with PDB40D-B

Method	Relative Time*	1% EPQ Cutoff	Coverage at 1% EPQ
SSEARCH % identity: within alignment	25.5	>70%	<0.1
SSEARCH % identity: within both	25.5	34%	3.0
SSEARCH % identity: HSSP-scaled	25.5	35% (HSSP + 9.8)	4.0
SSEARCH Smith–Waterman raw scores	25.5	142	10.5
SSEARCH E-values	25.5	0.03	18.4
FASTA ktup = 1 E-values	3.9	0.03	17.9
FASTA ktup = 2 E-values	1.4	0.03	16.7
WU-BLAST2 P-values	1.1	0.003	17.5
BLAST P-values	1.0	0.00016	14.8

*Times are from large database searches with genome proteins.

extent of errors. Second, SSEARCH, WU-BLAST2, and FASTA ktup = 1 perform best, though BLAST and FASTA ktup = 2 detect most of the relationships found by the best procedures and are appropriate for rapid initial searches.

The homologous proteins that are found by sequence comparison can be distinguished with high reliability from the huge number of unrelated pairs. However, even the best database searching procedures tested fail to find the large majority of distant evolutionary relationships at an acceptable error rate. Thus, if the procedures assessed here fail to find a reliable match, it does not imply that the sequence is unique; rather, it indicates that any relatives it might have are distant ones.**

**Additional and updated information about this work, including supplementary figures, may be found at http://sss.stanford.edu/sss/.

The authors are grateful to Drs. A. G. Murzin, M. Levitt, S. R. Eddy, and G. Mitchison for valuable discussion. S.E.B. was principally supported by a St. John's College (Cambridge, UK) Benefactors' Scholarship and by the American Friends of Cambridge University. S.E.B. dedicates his contribution to the memory of Rabbi Albert T. and Clara S. Bilgray.

1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410.
2. Altschul, S. F. & Gish, W. (1996) *Methods Enzymol.* **266**, 460–480.
3. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
4. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995) *J. Mol. Biol.* **247**, 536–540.
5. Brenner, S. E., Chothia, C., Hubbard, T. J. P. & Murzin, A. G. (1996) *Methods Enzymol.* **266**, 635–643.
6. Pearson, W. R. (1991) *Genomics* **11**, 635–650.
7. Pearson, W. R. (1995) *Protein Sci.* **4**, 1145–1160.
8. Smith, T. F. & Waterman, M. S. (1981) *J. Mol. Biol.* **147**, 195–197.
9. George, D. G., Hunt, L. T. & Barker, W. C. (1996) *Methods Enzymol.* **266**, 41–59.
10. Vogt, G., Etzold, T. & Argos, P. (1995) *J. Mol. Biol.* **249**, 816–831.
11. Henikoff, S. & Henikoff, J. G. (1993) *Proteins* **17**, 49–61.
12. Bairoch, A. & Apweiler, R. (1996) *Nucleic Acids Res.* **24**, 21–25.
13. Bairoch, A., Bucher, P. & Hofmann, K. (1996) *Nucleic Acids Res.* **24**, 189–196.
14. Henikoff, S. & Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919.
15. Dayhoff, M., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. (National Biomedical Research Foundation, Silver Spring, MD), Vol. 5, Suppl. 3, pp. 345–352.
16. Brenner, S. E. (1996) Ph.D. thesis. (University of Cambridge, UK).
17. Sander, C. & Schneider, R. (1991) *Proteins* **9**, 56–68.
18. Johnson, M. S. & Overington, J. P. (1993) *J. Mol. Biol.* **233**, 716–738.
19. Barton, G. J. & Sternberg, M. J. E. (1987) *Protein Eng.* **1**, 89–94.
20. Lesk, A. M., Levitt, M. & Chothia, C. (1986) *Protein Eng.* **1**, 77–78.
21. Arratia, R., Gordon, L. & M. W. (1986) *Ann. Stat.* **14**, 971–993.
22. Karlin, S. & Altschul, S. F. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2264–2268.
23. Karlin, S. & Altschul, S. F. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5873–5877.
24. Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. (1994) *Nat. Genet.* **6**, 119–129.
25. Pearson, W. R. (1996) *Methods Enzymol.* **266**, 227–258.
26. Lipman, D. J., Wilbur, W. J., Smith, T. F. & Waterman, M. S. (1984) *Nucleic Acids Res.* **12**, 215–226.
27. Wootton, J. C. & Federhen, S. (1996) *Methods Enzymol.* **266**, 554–571.
28. Waterman, M. S. & Vingron, M. (1994) *Stat. Science* **9**, 367–381.
29. Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) *J. Mol. Biol.* **13**, 669–678.
30. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987) in *Crystallographic Databases: Information Content, Software Systems, Scientific Applications*, eds. Allen, F. H., Bergerhoff, G. & Sievers, R. (Data Comm. Intl. Union Crystallogr., Cambridge, UK), pp. 107–132.
31. Brenner, S. E., Chothia, C. & Hubbard, T. J. P. (1997) *Curr. Opin. Struct. Biol.* **7**, 369–376.
32. Orengo, C., Michie, A., Jones S. Jones D. T. Swindells M. B. & Thornton, J. (1997) *Structure (London)* **5**, 1093–1108.
33. Zweig, M. H. & Campbell, G. (1993) *Clin. Chem.* **39**, 561–577.
34. Gribskov, M. & Robinson, N. L. (1996) *Comput. Chem.* **20**, 25–33.
35. Fitch, W. M. (1966) *J. Mol. Biol.* **16**, 9–16.
36. Chung, S. Y. & Subbiah, S. (1996) *Structure (London)* **4**, 1123–1127.
37. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
38. Girling, R., Schmidt, W., Jr, Houston, T., Amma, E. & Huisman, T. (1979) *J. Mol. Biol.* **131**, 417–433.
39. Spezio, M., Wilson, D. & Karplus, P. (1993) *Biochemistry* **32**, 9906–9916
40. Sayle, R. A. & Milner-White, E. J. (1995) *Trends Biochem. Sci.* **20**, 374–376.

Biochem. J. (1992) 288, 911-917 (Printed in Great Britain)

Cloning and functional expression of the canine anaphylatoxin C5a receptor

Evidence for high interspecies variability

Jason J. PERRET,* Eric RASPE,* Gilbert VASSART*† and Marc PARMENTIER*
*I.R.I.B.H.N., Faculté de Médecine, Université Libre de Bruxelles, Campus Erasme, Bat. C, 808 Route de Lennik, 1070 Bruxelles, Belgium, and †Service de Génétique Médicale, Hôpital Erasme, 808 route de Lennik, 1070 Bruxelles, Belgium

A cDNA clone, DTJP03, encoding an orphan receptor, was isolated from a canine thyroid library, and found to exhibit 68.6% amino-acid identity with the recently described human C5a receptor. This relatively low similarity first suggested that DTJP03 encoded either a C5a receptor subtype, or the presumably related C3a receptor. Binding studies performed on membranes from COS-7 cells expressing the recombinant receptor demonstrated that DTJP03 encoded a high-affinity C5a receptor, with a K_d of 1.2 nM. C3a was unable to compete for C5a binding. Intracellular free calcium concentrations were measured by Quin-2 fluorescence assays in Chinese hamster ovary cells stably transfected with the canine C5a receptor. C5a addition elicited an increase in the intracellular calcium concentration. Extracellular EGTA partially prevented this response, suggesting that activation of the C5a receptor promotes both the release of calcium from intracellular stores, and the influx of extracellular calcium. Genes encoding C5a-receptor subtypes were subsequently searched for by PCR in genomic DNA from human, canine, rat and bovine sources. The result was the amplification of a single gene fragment from each species, with about 70% identity between any two of them. The canine C5a receptor has therefore to be considered as orthologous to the human C5a receptor described previously. The low similarity between C5a receptors from different mammalian species is quite unusual for a G-protein-coupled receptor.

INTRODUCTION

In the course of complement activation, specific cleavage of the C5 component releases the anaphylatoxin C5a, a 74 amino-acid peptide (Fernandez & Hugli, 1978; Hugli, 1981). Binding of C5a to its specific membrane receptor induces physiological responses in a variety of cell types. *In vivo*, C5a is a potent mediator of the acute inflammatory response. *In vitro*, C5a exerts chemotaxis on macrophages and polymorphonuclear leucocytes and is a powerful stimulator of neutrophil function. It induces exocytosis of lysosomal hydrolytic enzymes, enhances production of superoxide radicals, promotes neutrophil leukotriene B_4 synthesis, as well as aggregation, adherence and neutrophil margination. In addition, C5a has spasmogenic effects, stimulating smooth muscle contraction, increasing vascular permeability and promoting mast-cell degranulation and histamine release. C5a also induces serotonin release from platelets, enhances interleukin-1 (IL-1) secretion from macrophages and up-regulates surface expression of the complement receptors CR1 and CR3 (for review see Fearon & Wong, 1983; Goldstein, 1988; Franck & Fries, 1991).

It was known that the C5a receptor belonged to the G-protein-coupled family, but different G-proteins and intracellular pathways seem to be involved in the signal transduction. Most of the effects induced by C5a are mediated through coupling of the C5a receptor to pertussis toxin (PT)-sensitive G-protein(s) (Warner *et al.*, 1987; Nourshargh & Williams, 1990; Rollins *et al.*, 1991). Some effects are, however, insensitive to PT treatment (Monk & Banks, 1991*a*,*b*). The requirement for extracellular calcium also depends on the assay system (Zimmerli *et al.*, 1990; Dore *et al.*, 1990; Kernen *et al.*, 1991).

Using degenerate primers corresponding to the conserved transmembrane segments of the G-protein-coupled-receptor superfamily, we have isolated by PCR (Saiki *et al.*, 1988) a series of orphan receptors from either cDNA or genomic DNA (Libert *et al.*, 1989; Parmentier *et al.*, 1989). Some of these orphan receptors have since been identified (Maenhaut *et al.*, 1990; Libert *et al.*, 1991; Parmentier *et al.*, 1992).

Recently, the cloning of the human C5a receptor was reported (Gerard & Gerard, 1991; Boulay *et al.*, 1991) and sequence comparison revealed a 68.6% identity with one of our orphan receptors, DTJP03. In this paper we present results confirming that DTJP03 encodes the canine C5a receptor. In the search of potential subtypes, partial clones encoding the canine, human, rat and bovine C5a receptors were amplified by PCR. A single type was obtained for each species, sharing about 70% identity with one another. This represents a surprisingly high interspecies variability as compared with other G-protein-coupled receptors.

MATERIALS AND METHODS

Cloning and sequencing

PCR was performed on 0.15 μg of purified cDNA as described previously (Libert *et al.*, 1989; Parmentier *et al.*, 1989), and the amplification product was cloned in M13 vectors for sequencing. JP03, a 600 bp fragment encoding part of an orphan receptor, was used to screen a canine thyroid λgt11 cDNA library (Lefort *et al.*, 1989). The two positive clones were purified to homogeneity and the *Eco*RI cDNA inserts were subcloned in pBluescript SK+ plasmid vector (Stratagene). After subcloning of overlapping restriction fragments in M13mp18 and 19, the clones were sequenced on both strands by the dideoxynucleotide-chain-termination method (Sanger *et al.*, 1977), using an automated DNA sequencer (Applied Biosystems 370A).

The coding region was cloned as a 1350 bp *Eco*RI–*Pst*I fragment in pBluescript SK+ and further subcloned as a

Abbreviations used: IL-1 interleukin 1; IL-8, interleukin 8; PT, pertussis toxin; CHO, Chinese hamster ovary; fMLP, formyl Met-Leu-Phe; ICL, intracellular loops.

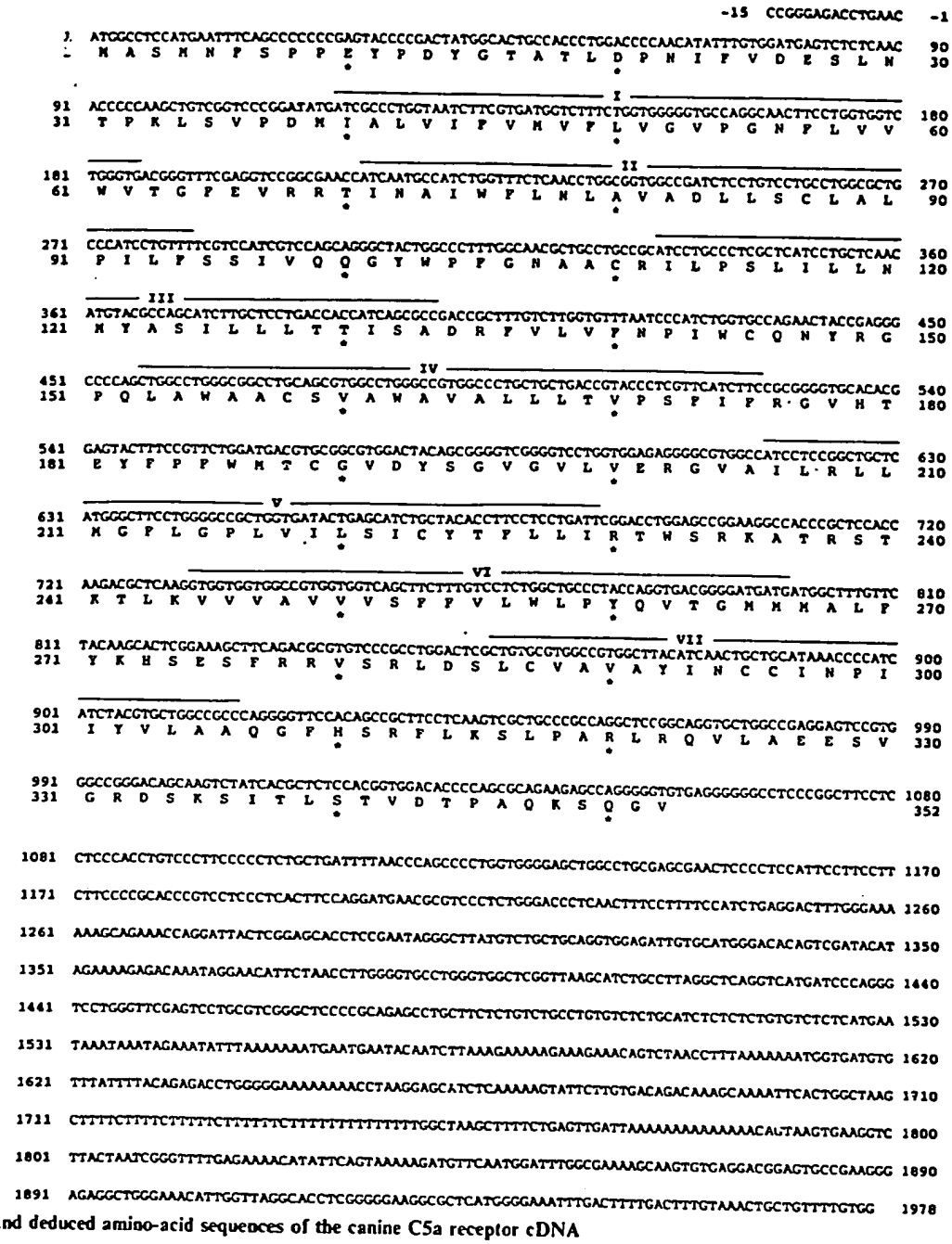


Fig. 1. Nucleotide and deduced amino-acid sequences of the canine C5a receptor cDNA

Numbering is relative to the putative initiation codon which presents a satisfactory sequence context according to Kozak (1989). The putative transmembrane segments are indicated by roman numerals I to VII. * indicates every tenth amino acid.

*Bam*HI–*Xho*I fragment in the eukaryotic expression vector pSVL (Pharmacia).

Sequence handling and data analysis were carried out using DNASIS/PROSIS software (Hitachi), LWL85 software (Li & Luo, 1985) and the GCG/VMS software package (Genetic Computer Group, WI, U.S.A.).

Transfections and cell cultures

The pSVL/DTJP03 construct was transfected in COS-7 cells as described previously (Gérard *et al.*, 1991). Chinese hamster ovary (CHO) CHO-K1 cells were co-transfected with pSVL/DTJP03 and pSV_2Neo as described (Perret *et al.*, 1990). After 10 days of G418 selection, resistant cells were pooled and stored in liquid nitrogen until needed. COS-7 and CHO-K1 cells were cultured using Dulbecco's modified Eagle's medium and Ham's F12 medium respectively, as described previously (Perret *et al.*, 1990). Clonal CHO cell lines were produced by low-density-culture seeding and harvesting of individual clones using sterile rings.

Membrane preparations

COS-7 cell membranes were prepared 72 h after transfection. CHO-K1 cells were prepared using nearly confluent cultures. Membrane preparation was as previously described (Perret *et al.*, 1990). Protein content was determined using the Bradford assay (Bradford, 1976).

Ligands

Human C5a, human C3a, ^{125}I-labelled human C5a and ^{125}I-labelled human C3a were kindly provided by Dr. Bitter-

Table 1. Nucleotide and amino-acid identity scores and synonymous (k_s) and non-synonymous (k_a) evolutionary rates for the C5a receptor and several other members of the G-protein-coupled-receptor family

K_a and k_a values calculated using LWL85 software (Li & Luo, 1985). Bovine and Rat C5a receptor sequences are partial sequences extending from transmembrane segments II to V and II to VII respectively.

Receptor	Species		Evolutionary rates		Identity (%)	
			k_s/year	k_a/year	Nucleotide	Amino acid
Complement C5A	HUM	DOG	2.7	1.3	75.2	68.6
	HUM	RAT	6.1	1.3	73.6	68.1
	HUM	BOV	3.0	0.9	80.5	76.4
	DOG	RAT	6.1	1.6	72.4	64.3
	DOG	BOV	2.6	1.3	78.9	68.2
	RAT	BOV	5.4	1.3	73.2	67.6
Tachykinin NK2	HUM	RAT	3.7	0.5	86.5	88.4
Dopamine D2	HUM	MUS	2.8	0.1	90.7	96.0
B2 Adrenergic	HUM	MUS	3.8	0.4	83.0	85.9
	HUM	RAT	3.3	0.4	81.0	86.4
TSH	HUM	RAT	3.9	0.5	84.4	85.7
	DOG	RAT	4.6	0.4	85.1	89.0
	DOG	HUM	2.5	0.4	89.8	89.7
Muscarinic M1	HUM	PIG	2.4	0.0	92.4	99.1
	HUM	RAT	2.6	0.0	91.4	98.7
	HUM	MUS	2.9	0.1	90.2	98.0
	MUS	PIG	4.0	0.1	88.0	97.2
	PIG	RAT	3.4	0.1	89.4	97.8

Suermann and coworkers (Institut für Medizinische Mikrobiologie, Medizinische Hochschule, Hannover, Germany).

Binding assays

All assays were carried out in 5-ml polypropylene tubes in a final volume of 100 µl, and incubated for 1 h with constant shaking at room temperature. The assay was initiated by the addition of membranes (75–100 µg of protein) to the tubes containing 20 mM-Hepes (pH 7.4)/125 mM-NaCl/5 mM-KCl/0.5 mM-glucose/0.25% (w/v) BSA/1 mM-$CaCl_2$/1 mM-$MgCl_2$ and the ligands. Concentration of labelled ligands in displacement experiments were 1–5 nM ^{125}I-labelled human C5a and 10 nM ^{125}I-labelled human C3a. Non-specific binding was determined by adding an excess of either human C5a (1 µM) or human C3a (10 µM). The assay was terminated by centrifuging (13000 *g*, 0 °C, 10 min) the membranes through a 10% (w/v) sucrose cushion in phosphate-buffered saline. The tubes were then frozen in liquid nitrogen for 2–3 min, and the bottom of each tube containing the membrane pellet was cut with a blade and counted in a gamma counter.

Calcium assay

CHO cell cultures, preparation for assay, and assay conditions were carried out as described previously (Van Sande *et al.*, 1990). Human C5a was used at a concentration of 100–150 nM. To assay the contribution of extracellular calcium, EGTA (1.5 mM) was added to the assay buffer in some experiments.

Genomic PCR

Aliquots (1 µg) of canine, human, rat and bovine genomic DNA were used as a target DNA in PCR reactions (30 cycles of 1 min at 93 °C, 2 min at 55 °C, 3 min at 72 °C). Other annealing temperatures were tested: 52 °C, 50 °C, 48 °C and 45 °C. All other conditions were as for PCR on the canine thyroid cDNA library described above. Primers containing *Xba*I or *Hin*dIII restriction sites for cloning were as follows:

P1 = TAGATCTAGATCAA(T/C)GC(G/C)AT(C/A/T)-TGGTT(T/C)CT;

P2 = ACTTAAGCTT(T/G)ATGCAGCA(G/A)TT(C/A/T)-AT(G/A)TA;

P3 = TAGATCTAGACTGTTTTCGTCCATCGTCCA;

P4 = ACTTAAGCTTACCAC(G/C)ACCTT(C/T)-AGTGT(C/T)TT.

RESULTS AND DISCUSSION

Numerous discrete bands were obtained in PCR reactions, using cDNA from a canine thyroid λgt11 library (Lefort *et al.*, 1989) as target DNA, and degenerate primers corresponding to the conserved regions of the second, third, and seventh transmembrane segments of known G-protein-coupled receptors, as described previously (Libert *et al.*, 1989; Parmentier *et al.*, 1989). These bands were cloned into the M13 vectors and sequenced. Open reading frames presenting similarities with G-protein-coupled receptors were searched for in all frames. Three clones encoding new putative members of the receptor superfamily were used as probes to screen the canine thyroid cDNA library. DTJP03, a 600 bp PCR clone, produced two positive signals out of 10^6 clones screened. The larger clone (1993 bp) contained entirely the smaller one (1.5 kb). Sequencing revealed a 1056 bp open reading frame (Fig. 1), having two potential AUG initiation codons (bases 1 and 10). The region surrounding the first AUG is closer to the consensus for initiation sites, as described by Kozak (1989). The coding sequence of DTJP03 was cloned as a 1350 bp insert into pBluescript SK+ and in the pSVL eukaryotic expression vector.

The cDNA encodes a 352 amino-acid protein (Fig. 1) with a

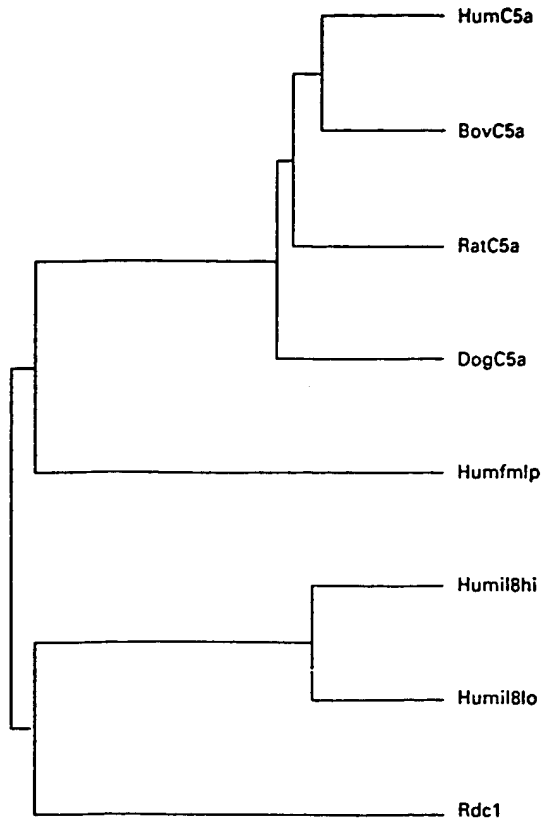


Fig. 2. Dendrogram showing the relative sequence similarities between the C5a receptor from four mammalian species, and the related receptors from the G-protein-coupled family

The dendrogram was generated by using the GCG Pileup software. Abbreviations: HumC5a, human C5a; BovC5a, bovine C5a; RatC5a, rat C5a; DogC5a, dog C5a; Humfmlp, human fMLP receptor; Humil8hi, human IL-8 high-affinity receptor; Humil8lo, human IL-8 low-affinity receptor; Rdc1, our orphan receptor (Libert *et al.*, 1989).

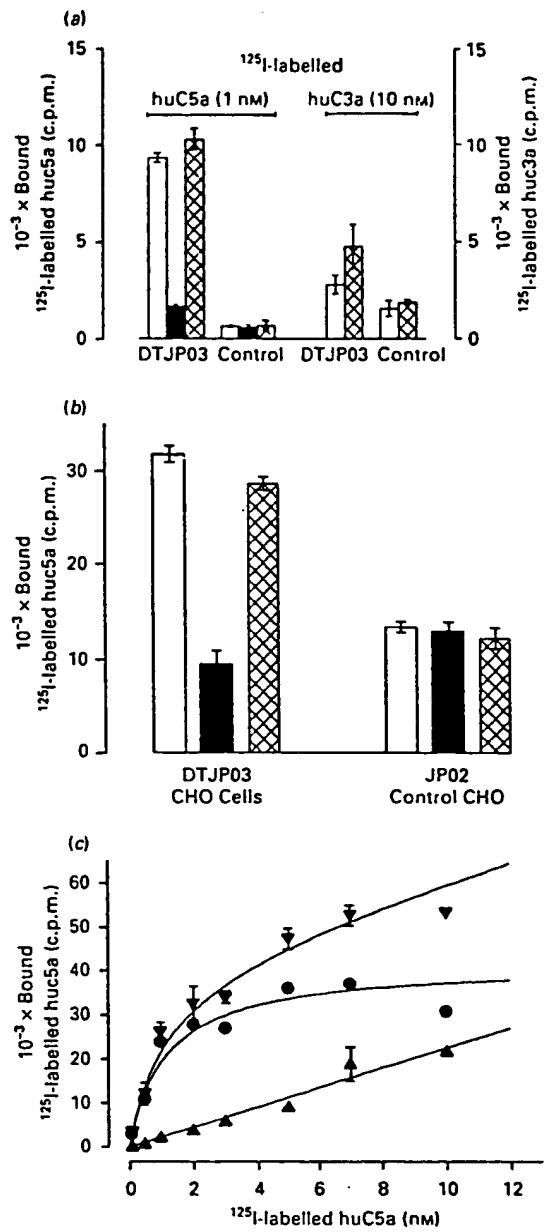


Fig. 3. Binding studies on the canine C5a receptor expressed in COS-7 and CHO cell lines

(*a*) Binding of ^{125}I-labelled human C5a (huC5a) and ^{125}I-labelled human C3a (huC3a) to COS-7 cells transiently expressing DTJP03, and to control COS-7 cells. Bound radiolabelled ligand (open bars) was displaced by an excess of either unlabelled huC5a (1 μM) (■) or huC3a (10 μM) (⊠). (*b*) A similar experiment was performed on CHO cells stably transfected with DTJP03, as compared with CHO cells transfected with the pSV$_2$Neo plasmid alone (JP02 control). (*c*) Saturation binding experiment performed on COS-7 cells transiently expressing the canine C5a receptor. Total (▼), specific (●) and non-specific binding (▲) are represented. The non-specific binding was determined in the presence of 1 μM unlabelled huC5a. Curve fitting using a non-linear regression algorithm yielded an apparent K_d of 1.2 nM.

calculated relative molecular mass of 39186. The hydropathy profile (Kyte & Doolittle, 1982) of the deduced amino-acid sequence is consistent with the presence of seven transmembrane domains (results not shown). Sequence comparison with the recently cloned human C5a receptor (Gerard & Gerard, 1991; Boulay *et al.*, 1991) revealed a 68% amino-acid identity. This percentage is well below the identity scores obtained between mammalian orthologues for other G-protein-coupled receptors, which generally ranged between 85 and 98%. A few examples are given for comparison in Table 1. From the dendrogram displayed in Fig. 2, the similarity is of the same order as that observed for the two recently cloned high- and low-affinity interleukin-8 (IL-8) receptors (Holmes *et al.*, 1991; Murphy & Tiffany, 1991). Given this moderate similarity, our first hypothesis was that DTJP03 encoded a receptor closely related to, but different from the published human C5a receptor. We considered that it could encode either a C5a subtype, or the presumably related C3a receptor as suggested by the similar structure of the two ligands (Greer, 1986).

In order to assay the binding characteristics of DTJP03, membranes were prepared from transfected COS-7 cells transiently expressing the recombinant receptor. Transfected COS membranes exhibited a high binding capacity for ^{125}I-labelled human C5a (8000 c.p.m.). Over 80% of the total binding was specifically displaced by a 10^3-fold excess of unlabelled C5a, a 10^4-fold excess of unlabelled C3a being without effect (Fig. 3*a*). No specific binding of ^{125}I-labelled human C3a was obtained with transfected cells. Control COS cell membranes did not display significant binding for either C5a (700 c.p.m.) or C3a. Saturation curves obtained with transfected COS cell membranes and increasing concentrations of ^{125}I-labelled human C5a demonstrated a saturable high-affinity binding site with an apparent K_d of 1.2 nM (Fig. 3*c*), this being within the range

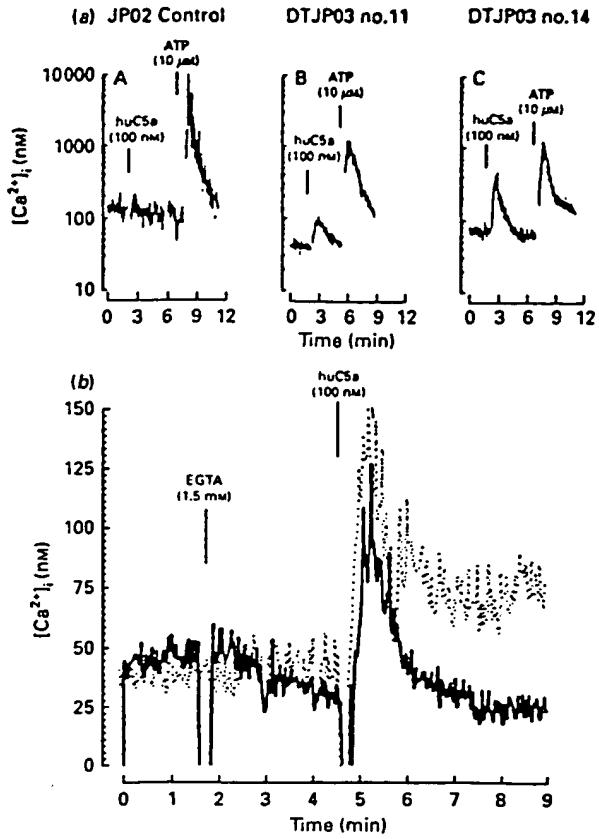


Fig. 4. Intracellular calcium measurements in stably transfected CHO cells and JP02 control CHO cells

Experiments were performed using a Quin-2 fluorescence assay and a SPEX fluorimeter. (*a*) Panels A, B and C represent respectively the responses elicited by the successive addition of human C5a (huC5a) (100 nM) and ATP (10 μM) on a CHO cell line transfected with pSV_2Neo (JP02 control), and two clonal cell lines (nos. 11 and 14) transfected with DTJP03, and expressing the C5a receptor with a different efficiency. (*b*) Effect of the extracellular calcium concentration on the response to C5a in clone no. 14. The dotted line represents the response to huC5a (100 nM) in control medium containing 1 mM-Ca^{2+}. The continuous line represents the response to huC5a after addition of an extracellular excess of EGTA (1.5 mM).

previously described for the human neutrophil and macrophage C5a receptors (Chenoweth *et al.*, 1982; Huey & Hugli, 1985; Yancey *et al.*, 1989).

CHO-K1 cells were co-transfected by the pSVL/DTJP03 construct and the pSV_2Neo vector conferring resistance to neomycin. After selection by G418, the pool of resistant clones was used to prepare membranes. Specific binding of ^{125}I-labelled human C5a to stably transfected CHO cell-membrane preparations yielded results similar to those obtained with COS-7 cells (Fig. 3*b*), while C3a was unable to bind to the receptor. The pool of neomycin-resistant CHO cells was cloned by high-dilution-culture seeding followed by subsequent recovery of the isolated colonies. The individual clones were screened by measuring intracellular calcium concentrations by the Quin-2 fluorescence assay (Van Sande *et al.*, 1990). Only 25% (6/24) of the clones responded to 150 nM-human C5a in the Ca^{2+} assay. One clone (No. 14) presented a large response to C5a, as compared with that elicited by 10 μM-ATP, used as positive control (Fig. 4*a*). This clone was used for further studies. Addition of 1.5 mM-EGTA to the extracellular assay medium modified the calcium fluorescence signal induced by subsequent addition of the agonist (Fig. 4*b*). In the presence of extracellular calcium, the intracellular calcium concentration increased rapidly to its maximal level within 1 min, then decreased gradually back to the basal level or slightly higher. However, when extracellular calcium was chelated by excess EGTA, human C5a evoked a sharp transient increase of intracellular calcium, that rapidly fell off to beneath the basal level. As the initial increase in calcium concentration is resistant to extracellular calcium depletion by EGTA, it is likely that calcium is released from intracellular stores. On the other hand, the subsequent decrease in calcium concentration is faster in the absence of extracellular calcium, suggesting that calcium influx contributes to the sustained phase. This behaviour is typical of receptors coupled to a G-protein activating the inositol phosphate cascade (Kojima, 1990; Dore *et al.*, 1990).

These results clearly demonstrate the DTJP03 is a canine high-affinity C5a receptor that functionally couples to the $InsP_3$–calcium cascade in stably transfected CHO cell lines. Although there is no pharmacological evidence for C5a receptor subtypes, the low degree of similarity between the canine and the human receptors could be indicative of a possible molecular heterogeneity of the C5a receptor. We therefore searched for orthologues of our canine C5a and of the human C5a receptors in four mammalian species: human, dog, rat and cow. A PCR-based approach was used, in which four nucleotide sequences, conserved between the cloned canine and human C5a receptors, were used as primers to amplify related gene fragments from genomic DNA. Like most of the other members of the seven transmembrane G-protein-coupled-receptor superfamily, the C5a receptor lacks introns in its coding sequence. Four moderately degenerate primers corresponding to parts of transmembrane segments II (P1 and P2), VI (P3) and VII (P4) were defined (see the Materials and methods section). Independent PCR reactions were performed, using all four primer combinations (i.e. P1 versus P2; P1 versus P4; P3 versus P2 and P3 versus P4). To allow primer hybridization in the presence of potential mismatches, several annealing temperatures (55 °C, 52 °C, 50 °C, 48 °C, and 45 °C) were used. Down to 45 °C, only one band was visible for each primer combination at the expected size (results not shown). These bands were cloned into the bacteriophage M13mp18 and 19 vectors and sequences were obtained from five or six clones under each condition. All sequences obtained from canine DNA were identical to DTJP03; likewise all sequences from human genomic DNA were identical to the published human C5a receptor sequence. The bovine and rat sequences were unique as well, regardless of PCR stringency or primer combination. The partial amino-acid sequences obtained from rat and bovine sequences were aligned with the canine and human sequences (Fig. 5). Similarity between any two sequences was close to 70%. These results demonstrate that our canine receptor and the reported human C5a receptors effectively represent orthologues.

There remains then the question of the surprisingly low interspecies amino-acid conservation (68.9 ± 3.7%), which is in contrast to the high conservation between other G-protein-coupled receptors, such as the cannabinoid, the thyroid stimulating hormone or the muscarinic receptors. This contrast appears clearly when the evolutionary rates for synonymous changes (k_s) and non-synonymous changes (k_a) (Li & Luo, 1985) are calculated for the C5a receptors and several other members of the G-protein-coupled-receptor family, as well as the nucleotide and amino-acid identities (Table 1). As expected, the k_s values, reflecting the rate of nucleotide substitutions that do not affect amino acids, are relatively constant for all receptors, as it reflects solely the evolutionary distance between species. On the other hand, the k_a values, reflecting the nucleotide substitution rate affecting the amino-acid sequence, are low for most

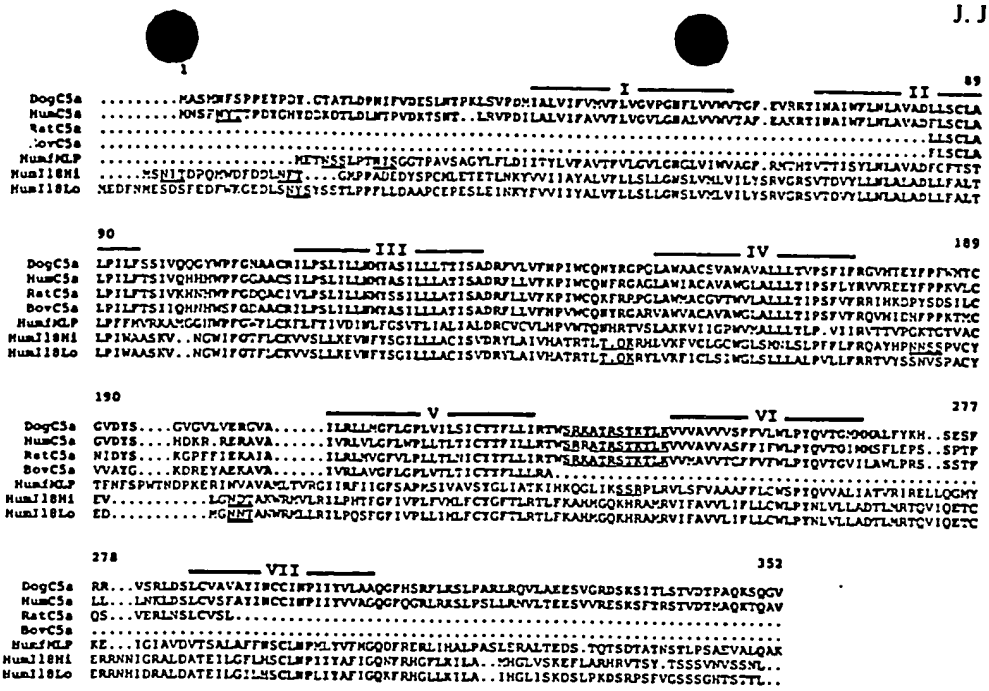


Fig. 5. Amino-acid alignment of the canine and human C5a receptors, the partial rat and bovine sequences obtained from genomic DNA by PCR, and the sequences from the related chemoattractant receptors to human IL-8 and human fMLP

Amino acids identical to the canine C5a receptor are represented in bold characters. Numbering is relative to the canine C5a receptor amino-acid sequence. The transmembrane segments are overlined and numbered I to VII. The putative sites for *N*-linked glycosylation in the *N*-terminal and extracellular domains are underlined twice; the putative sites for phosphorylation by protein kinases A and C in the second and third ICLs are underlined once. The alignment was performed by using the GCG Pileup software. See legend to Fig. 2 for key to abbreviations.

G-protein-coupled receptors, and significantly higher for the C5a receptor. A corollary of these high k_a values is the lower identity at the amino-acid level as compared with the nucleotide level in contrast to the increase observed for the other members of the G-protein-coupled family. This signifies that evolutionary constraints affect quite differently classes of receptors that are believed to share a common transmembrane organization and structure–function relationships. It is unclear to us why the C5a receptor appears to deviate from the strong conservation prevailing in this gene family.

Northern-blot analysis (results not shown) on different canine tissues detected C5a receptor transcripts in thyroid (but not in cultured thyrocytes), testis, brain, lung, kidney, spleen and stomach. This partial distribution, as well as the cloning from a thyroid library, reflects the presence of circulating leucocytes and tissue macrophages in all organs.

Within this background of high interspecies variability, the similarities between protein sequences could pinpoint functionally important conserved segments or residues, implicated either in ligand–receptor interactions, G-protein-coupling or receptor desensitization. We therefore aligned the C5a receptor sequences with their closest relatives of the G-protein-coupled family, the human formyl Met-Leu-Phe (fMLP) (Boulay *et al.*, 1990) and the two IL-8 receptors (Fig. 5). The extracellular *N*-terminal domain is poorly conserved between the canine and human receptors with only 44% (17/39) identical residues. It has been proposed that the high negative charges of the receptor are important for the interaction with the positively charged ligand, although few of these charges are on the surface of the C5a ligand (Zuiderweg *et al.*, 1989; Mollison *et al.*, 1989). Within the *N*-terminal domain of the receptor, only three charged residues are identical in both species, and the same observation prevails for the extracellular loops that are the least conserved parts of the receptor. This poor conservation, together with the absence of species selectivity for the ligand (the canine receptor interacts very efficiently with human C5a), do not give support to the involvement of the extracellular domains in receptor–ligand interactions.

Interestingly, in contrast to the human C5a receptor, the canine receptor does not present potential *N*-glycosylation sites in its extracellular *N*-terminus. The conserved Asn-Xaa-Thr sequence is, in the canine sequence, followed by a proline, which has been shown to prevent glycosylation (Pless & Lennarz, 1977; Bause, 1983). Such differential glycosylation is also observed for the C5a ligand which is glycosylated in human (Zuiderweg *et al.*, 1989) but not in pig (Williamson & Madison, 1990). The intracellular *C*-terminus contains numerous serine and threonine residues (10 for canine; 11 for human) that are potential phosphorylation sites for β-ARK-related serine/threonine protein kinases (Benovic *et al.*, 1989). Another interesting feature of the *C*-terminus, shared with the bovine NPY (Rimland *et al.*, 1991), human IL-8, RDC1 (Libert *et al.*, 1989), and the human fMLP receptors, is the absence of a cysteine residue, conserved among most G-protein-coupled receptors, thought to serve as a site for palmitoylation (Dohlman *et al.*, 1991).

Potential phosphorylation sites, conserved in the four species, are present in the third intracellular loop: one for the cyclic AMP-dependent kinases (protein kinase A) (Feramisco *et al.*, 1980; Glass *et al.*, 1986) at residue Thr-237, and three for the protein kinase C (Woodgett *et al.*, 1986) at residues Ser-233, Ser-239 and Thr-242. Conserved residues in the C5a receptors include Phe-136 which replaces the tyrosine residue of the Asp-Arg-Tyr tri-peptide motif (end of TM3) common to most G-protein-coupled receptors.

The transmembrane segments are the most conserved, with 73% identity (119/162); however, the intracellular loops (ICL) are also highly conserved with 70% (31/44) of identical residues. The lowest identity is observed for ICL2 (58%, 11/19). ICL3 presents up to 87% (13/15) amino-acid identity within the C5a receptor group. The C5a, fMLP and IL-8 receptors share very little similarity within ICL3, despite the coupling of these receptors to a common intracellular pathway, and the involve-

ment of ICL3 in G-protein coupling (Dohlman *et al.*, 1991; Bonner, 1992). Charged residues thought to interact with negative charges of G-proteins (Dohlman *et al.*, 1991) are, however, present in all cases.

As a conclusion, cloning of the C5a receptors in several species along with the characterization of the corresponding ligands should allow a better approach in determining critical amino acids implicated in receptor–ligand interaction, G-protein-coupling and desensitization. It will pinpoint candidate residues for mutagenesis and chimeric constructions, both in the receptors and the ligands. Ultimately, the availability of the cloned receptors should help the design of pharmacologically active (non-peptide) inhibitors that could be used in syndromes were inappropriate complement activation occurs.

We would like to thank Drs. J. E. Dumont, J. Khôl and A. Klos for useful discussions, Dr. S. Swillens for his help in binding studies and computer analysis of results, Dr. P. Cochaux and C. Mollereau for lending a hand when most needed. We also appreciated the technical help of Y. Mauquoy in maintaining the cell lines. This work was supported by the Belgian programme on Inter-University Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming. It was also supported by the Fonds de la Recherche Scientifique Médicale of Belgium, Boehringer Ingelheim and the Association Recherche Biomédicale et Diagnostic. The scientific responsibility is assumed by the authors. J.P. and E.R. are respectively fellows of the IRSIA and Televie. M.P. is Chercheur Qualifié of the Fonds National de la Recherche Scientifique.

REFERENCES

Bause, E. (1983) Biochem. J. 209, 331–336
Benovic, J. L., DeBlasi, A., Stone, W. C., Caron, M. G. & Lefkowitz, R. J. (1989) Science 246, 235–346
Bonner, T. I. (1992) Trends Pharmacol. Sci. 13, 48–50
Boulay, F., Tardif, M., Brouchon, L. & Vignais, P. (1990) Biochem. Biophys. Res. Commun. 168, 1103–1109
Boulay, F., Mery, L., Tardif, M., Brouchon, L. & Vignais, P. (1991) Biochemistry 30, 2983–2990
Bradford, M. (1976) Anal. Biochem. 72, 248–259
Chenoweth, D. E., Goodman, M. G. & Weigle, W. O. (1982) J. Exp. Med. 156, 68–78
Dohlman, H. G., Thorner, J., Caron, M. G. & Lefkowitz, R. J. (1991) Annu. Rev. Biochem. 60, 653–688
Dore, M., Slauson, D. O., Suyemoto, M. M. & Nielson, N. R. (1990) Inflammation 14, 71–82
Fearon, D. T. & Wong, W. W. (1983) Annu. Rev. Immunol. 1, 243–271
Feramisco, J. R., Glass, D. B. & Krebs, E. G. (1980) J. Biol. Chem. 255, 4240–4245
Fernandez, H. N. & Hugli, T. E. (1978) J. Biol. Chem. 253, 6955–6964
Franck, M. M. & Fries, L. F. (1991) Immunology Today 12, 322–326
Gerard, N. P. & Gerard, C. (1991) Nature (London) 349, 614–617
Gérard, C. M., Mollereau, C., Vassart, G. & Parmentier, M. (1991) Biochem. J. 279, 129–134
Glass, D. B., El-Maghrabi, M. R. & Pilkis, S. J. (1986) J. Biol. Chem. 261, 2987–2993
Goldstein, I. M. (1988) in Inflammation: Basic Principles and Clinical Correlates (Gallin, J. I., Goldstein, I. M. & Synderman, R., eds.), pp. 309–323, Raven Press, New York
Greer, J. (1986) Enzyme 36, 150–163
Holmes, W. E., Lee, J., Kuang, W. J., Rice, G. C. & Wood, W. I. (1991) Science 253, 1278–1280
Hucy, R. & Hugli, T. E. (1985) Immunology 235, 2063–2068
Hugli, T. E. (1981) CRC Crit. Rev. Immunol. 1, 321–366
Kernen, P., Wymann, M. P., von Tscharner, V., Deranleau, D. A., Tai, P. C., Spry, C. J., Dahinden, C. A. & Baggiolini, M. (1991) J. Clin. Invest. 87, 2012–2017
Kojima, I. (1990) in G Proteins (Iyengor, R. & Birnbaumer, L., eds.), chapt. 21, pp. 503–517, Academic Press, New York
Kozak, M. (1989) J. Cell Biol. 108, 229–241
Kyte, J. & Doolittle, R. F. (1982) J. Mol. Biol. 157, 105–132
Lefort, A., Lecocq, R., Libert, F., Lamy, F., Swillens, S., Vassart, G. & Dumont, J. E. (1989) EMBO J. 8, 111–116
Li, W. H. & Luo, C. C. (1985) Mol. Biol. Evol. 2, 150–174
Libert, F., Parmentier, M., Lefort, A., Dinsart, C., Van Sande, J., Maenhaut, C., Simons, M.-J., Dumont, J. E. & Vassart, G. (1989) Science 244, 569–572
Libert, F., Schiffmann, S., Lefort, A., Parmentier, M., Gérard, C. M., Dumont, J. E., Vanderhaeghan, J.-J. & Vassart, G. (1991) EMBO J. 10, 1677–1682
Maenhaut, C., Van Sande, J., Libert, F., Abramowicz, M., Parmentier, M., Vanderhaeghen, J.-J., Dumont, J. E., Vassart, G. & Schiffmann, S. (1990) Biochem. Biophys. Res. Commun. 173, 1169–1178
Mollison, K. W., Mandecki, W., Zuiderweg, E. R. P., Fayer, L., Fey, T. A., Krause, R. A., Conway, R. G., Miller, L., Edalji, R. P., Shallcross, M. A., Lane, B., Lawrence-Fox, J., Greer, J. & Carter, G. W. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 292–296
Monk, P. N. & Banks, P. (1991*a*) Biochim. Biophys. Acta 1092, 251–255
Monk, P. N. & Banks, P. (1991*b*) J. Mol. Endocrinol. 6, 241–247
Murphy, P. M. & Tiffany, H. L. (1991) Science 253, 1280–1283
Nourshargh, S. & Williams, T. J. (1990) J. Immunol. 145, 2633–2638
Parmentier, M., Libert, F., Maenhaut, C., Lefort, A., Gérard, C. M., Perret, J., Van Sande, J., Dumont, J. E. & Vassart, G. (1989) Science 246, 1620–1622
Parmentier, M., Libert, F., Schurmans, S., Schiffmann, S., Lefort, A., Eggerickx, D., Ledent, C., Mollereau, C., Gérard, C. M., Perret, J., Grootegoed, A. & Vassart, G. (1992) Nature (London) 355, 453–455
Perret, J., Ludgate, M., Libert, F., Gérard, C., Dumont, J. E., Vassart, G. & Parmentier, M. (1990) Biochem. Biophys. Res. Commun. 17, 1044–1050
Pless, D. D. & Lennarz, W. J. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 134–138
Rimland, J., Xin, W., Sweetman, P., Saijoh, K., Nestler, E. J. & Duman, R. S. (1991) Mol. Pharmacol. 40, 869–875
Rollins, T. E., Siciliano, S., Kobayashi, S., Cianciarulo, D. N., Bonilla-Argudo, V., Collier, K. & Springer, M. S. (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 971–975
Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) Science 239, 487–491
Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463–5467
Van Sande, J., Raspé, E., Perret, J., Lejeune, C., Maenhaut, C., Vassart, G. & Dumont, J. E. (1990) Mol. Cell. Endocrinol. 74, R1–R6
Warner, J. A., Yancey, K. B. & MacGlashan, D. W. (1987) J. Immunol. 139, 161–165
Williamson, M. P. & Madison, V. S. (1990) Biochemistry 29, 2895–2905
Woodgett, J. R., Gould, K. L. & Hunter, T. (1986) Eur. J. Biochem. 161, 177–184
Yancey, K. B., Lawley, T. J., Derssokian, M. & Harvath, L. (1989) J. Invest. Dermatol. 92, 184–189
Zimmerli, W., Reber, A. M. & Dahinden, C. A. (1990) J. Infect. Dis. 161, 242–249
Zuiderweg, E. R. P., Nettesheim, D. G., Mollison, K. W. & Carter, G. W. (1989) Biochemistry 28, 172–185

Received 5 June 1992; accepted 22 July 1992

Docket No.: PF-0040 US
USSN: 08/462,355
Ref. No. 3

Communication

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 34, Issue of August 23, pp. 20231–20234, 1996

Molecular Cloning and Characterization of the Human Anaphylatoxin C3a Receptor*

(Received for publication, June 18, 1996, and in revised form, June 26, 1996)

Robert S. Ames‡§, Yi Li¶, Henry M. Sarau‖, Paru Nuthulaganti, James J. Foley‖, Catherine Ellis**, Zhizhen Zeng¶, Kui Su¶, Anthony J. Jurewicz‡‡, Robert P. Hertzberg‡‡, Derk J. Bergsma**, and Chandrika Kumar****

*From the Departments of ‡Molecular Immunology, **Molecular Genetics, ‡‡Biomolecular Discovery, and ‖Pulmonary Pharmacology, SmithKline Beecham Pharmaceuticals, King of Prussia, Pennsylvania 19406-0939 and ¶Human Genome Sciences, Rockville, Maryland 20850-3338*

In a human neutrophil cDNA library, an orphan G-protein-coupled receptor, HNFAG09, with 37% nucleotide identity to the C5a receptor (C5a-R, CD88) was identified. A novel feature of this gene, unlike C5a-R and other G-protein-coupled receptors, is the presence of an extraordinarily large predicted extracellular loop comprised of in excess of 160 amino acid residues between transmembrane domains 4 and 5. Northern blot analysis revealed that expression of mRNA for this receptor in human tissues, while similar, was distinct from C5a-R expression. Although there were differences in expression, transcripts for both receptors were detected in tissues throughout the body and the central nervous system. Mammalian cells stably expressing HNFAG09 specifically bound ^{125}I-C3a and responded to a C3a carboxyl-terminal analogue synthetic peptide and to human C3a but not to rC5a with a robust calcium mobilization response. HNFAG09 encodes the human anaphylatoxin C3a receptor.

During complement activation the 74–77-amino acid anaphylatoxins C3a, C4a, and C5a are released. They are potent inflammatory mediators, inducing cellular degranulation, smooth muscle contraction, arachidonic acid metabolism, cytokine release, and cellular chemotaxis (reviewed in Refs. 1–3), and have been implicated in the pathogenesis of a number of inflammatory diseases (4, 5).

Studies have demonstrated the presence of a C3a receptor (C3a-R) on guinea pig platelets, rat mast cells, human neutrophils, eosinophils, and platelets (3). A single class of high affinity C3a binding sites has been characterized on human neutrophils and differentiated U937 cells (6). Competition binding and functional desensitization studies are consistent with the presence of a receptor for C3a, which is distinct from the C5a-R (3, 6). However, there is evidence that C3a and C4a may bind to the same receptor as the two anaphylatoxins cross-desensitize guinea pig ileal tissue (2, 3), although other investigators using guinea pig macrophages indicate that there may be separate receptors (7). Functional activity of the C3a-R is sensitive to pertussis toxin, consistent with the binding site being composed of a G-protein-coupled receptor (6).

A complete understanding of the role of C3a in the pathogenesis of the inflammatory response has been hampered by the lack of the cloned receptor. In this report we describe the molecular cloning, stable expression in mammalian cells, and functional characterization of the human C3a receptor. This same receptor was recently independently cloned from an HL-60 library by low stringency screening with a fMet-Leu-Phe receptor probe and, lacking functional data, claimed to be an orphan receptor (AZ3B) (8). Mouse L cells expressing AZ3B failed to bind and respond to the agonists examined, although C3a was not tested (8).

EXPERIMENTAL PROCEDURES

Materials—The C3a carboxyl-terminal analogue synthetic peptide (WWGKKYRASKLGLAR) (9) was obtained from Bachem Bioscience, Inc., King of Prussia, PA. C3a was purchased from Advanced Research Technologies, San Diego, CA. Human rC5a was expressed in *Escherichia coli* and purified to homogeneity. Other agonists were obtained from Sigma.

cDNA Cloning—cDNA library construction and screening were carried out essentially as described (10), and DNA sequence was determined using a ABI sequencer (11). Expressed sequence tag analysis (11–13) of cDNA clones derived from a human neutrophil (lipopolysaccharide activated) cDNA library (oligo(dT)-primed and constructed in the λ Uni-ZAP XR vector (Stratagene)) identified a clone demonstrating significant homology (approximately 40% amino acid sequence identity) to the C5a-R (14, 15). This cDNA clone contained an incomplete open reading frame and therefore was used to reprobe the neutrophil cDNA library to obtain a full-length cDNA. The alignment of HNFAG09 and the C5a-R was determined by the method of Needleman and Wunsch (21) using the Gap comparison program of the Wisconsin Package, version 8, September 1994, Genetics Computer Group, Madison, WI.

Northern Blot Analysis—Commercially prepared (Clontech, Palo Alto, CA) multiple tissue blots containing approximately 2 μg of poly(A) mRNA per lane were sequentially hybridized with random primer ^{32}P-labeled cDNAs spanning the coding regions of C5a-R and HNFAG09. C5a-R was cloned via PCR[1] from differentiated U937 RNA. The final washing step was carried out twice in 0.5 × SSC, 1% SDS at 65 °C for 20 min.

Stable Expression in RBL-2H3 Cells—To prepare HNFAG09 for expression in mammalian cells, a 1.6-kb cDNA fragment was obtained by PCR amplification that encompassed the entire HNFAG09 open reading frame. This fragment was subcloned into *KpnI*/*Hin*dIII sites of the mammalian expression vector, pCDN (16). Oligonucleotide primers used for PCR amplification were 5′-GA AGT GGT ACC ATG GCG TC -3′ and 5′- GC TCC AAG CTT TCA CAC AGT TG -3′ (the translation start and stop codons are underlined). RBL-2H3 cells were electroporated with either HNFAG09 or C5a-R in the pCDN mammalian expression vector (16), exactly as described (17). Individual G418-resistant (400 μg/ml) colonies were isolated and expanded. Clonal cell lines expressing either HNFAG09 or C5a-R were chosen for further functional and binding studies.

Calcium Mobilization—Fura 2-loaded clonal cell lines expressing C5a-R or HNFAG09 were assayed for functional response, Ca^{2+} mobilization, as described (18).

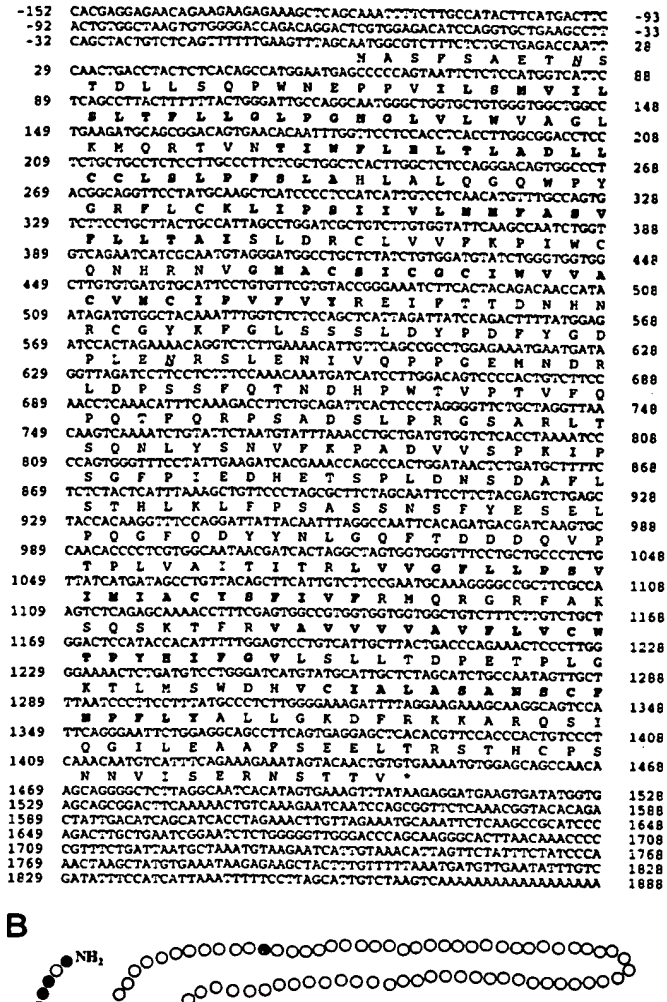
Binding Assay—C3a was radioiodinated using IODO-BEADS

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ To whom correspondence should be addressed: Dept. of Molecular Immunology UE0548, SmithKline Beecham Pharmaceuticals, 709 Swedeland Rd., P. O. Box 1539, King of Prussia, PA 19406-0939. Tel.: 610-270-7602; Fax: 610-270-5093.

[1] The abbreviations used are: PCR, polymerase chain reaction; kb, kilobase; PBL, peripheral blood leukocytes.

A



B

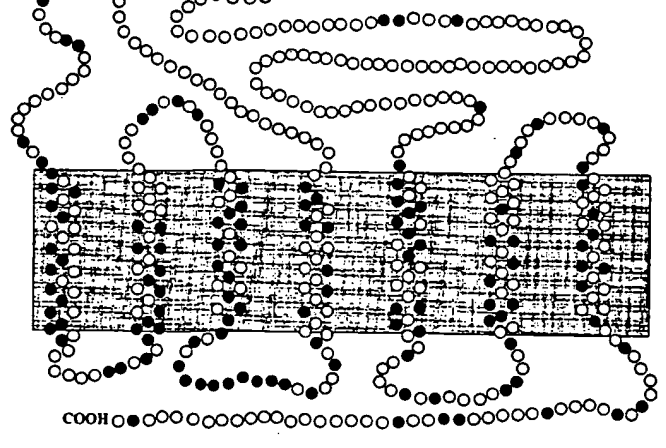


FIG. 1. *A*, nucleotide and deduced amino acid sequence of HNFAG09. The predicted seven-membrane spanning domains of HNFAG09 are indicated by *bold*, and glycosylation sites are indicated by *italics* and *underline*. This nucleotide sequence has been submitted to GenBank; the accession number is U62027. *B*, predicted membrane topology of HNFAG09. Amino acid residues in common between C5a-R and HNFAG09 have been highlighted *black*; two predicted *N*-linked glycosylation sites, in the large extracellular loop and the amino terminus, are indicated by *gray shading*.

(Pierce) to a specific activity of 100 Ci/mmol. Increasing concentrations of cold competitor were added to 1×10^6 cells in the presence of ^{125}I-C3a (2.3 nM), and the assay was performed essentially as described (6).

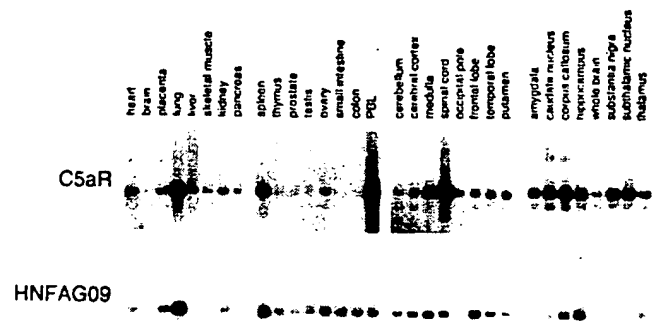


FIG. 2. **C5a-R and HNFAG09 transcripts are abundantly expressed in the central nervous system and throughout the body.** Tissue distribution of C5a-R and HNFAG09 as determined by Northern blot analysis. The tissue source of RNA is indicated above each *lane*.

RESULTS AND DISCUSSION

Expressed sequence tag analysis (11–13) of cDNA clones derived from a human neutrophil (lipopolysaccharide activated) cDNA library identified a clone demonstrating significant homology (approximately 40% amino acid sequence identity) to the C5a-R. This expressed sequence tag contained an incomplete open reading frame that therefore was used to reprobe the neutrophil cDNA library to obtain a 2040-base pair cDNA encoding a complete orphan G-protein-coupled receptor of 482 amino acids, which shared 37% nucleotide identity throughout the coding regions with the C5a-R (Fig. 1*A*). Although similar to the C5a-R, this cDNA contains two predicted extracellular *N*-linked glycosylation sites and an unusually large extracellular domain between transmembrane domains 4 and 5 comprised of over 160 amino acid residues (Fig. 1*A*). The majority of the identical residues between the C5a-R and HNFAG09 reside in the predicted transmembrane spanning domains and in the second intracellular loop (Fig. 1*B*).

By Northern blot analysis, expression of HNFAG09 in human tissues and cell lines is distinct from C5a-R expression. An ~2.2-kb C5a-R transcript was abundantly expressed in peripheral blood leukocytes (PBL), lung, spleen, heart, placenta, spinal cord, and throughout the brain. An ~2.1-kb HNFAG09 transcript was predominantly expressed in lung, spleen, ovary, placenta, small intestine, throughout the brain, and to a much lesser extent than C5a-R, in heart and PBL (Fig. 2). Although by Northern blot analysis the specific cells within the various tissues examined, which are expressing C5a-R and HNFAG09, cannot be determined, these data are suggestive that these receptors are abundantly expressed throughout the body. By fluorescent activated cell sorting using polyclonal antibodies generated to fusion proteins composed of glutathione *S*-transferase or maltose binding protein and the extracellular loop, this receptor has been shown to be expressed on several cell types, including U937, HL-60, PBL, and human neutrophils and monocytes (8).

Preliminary functional characterization in *Xenopus laevis* oocytes suggested that HNFAG09 encoded a human anaphylatoxin receptor.[2] To confirm these results in mammalian cells, this receptor was expressed in RBL-2H3 cells (19), a rat basophil cell line, which when transfected with an expression plasmid encoding the C5a-R expresses receptors that are functionally active (17). RBL-2H3 cells were stably transfected with mammalian expression plasmids encoding the C5a-R or HNFAG09, and Fura 2-loaded cells were tested for a C5a- or C3a-induced mobilization of intracellular Ca^{2+}. C5a-R- but not

[2] R. S. Ames, P. Nuthulaganti, and C. Kumar, unpublished observation.

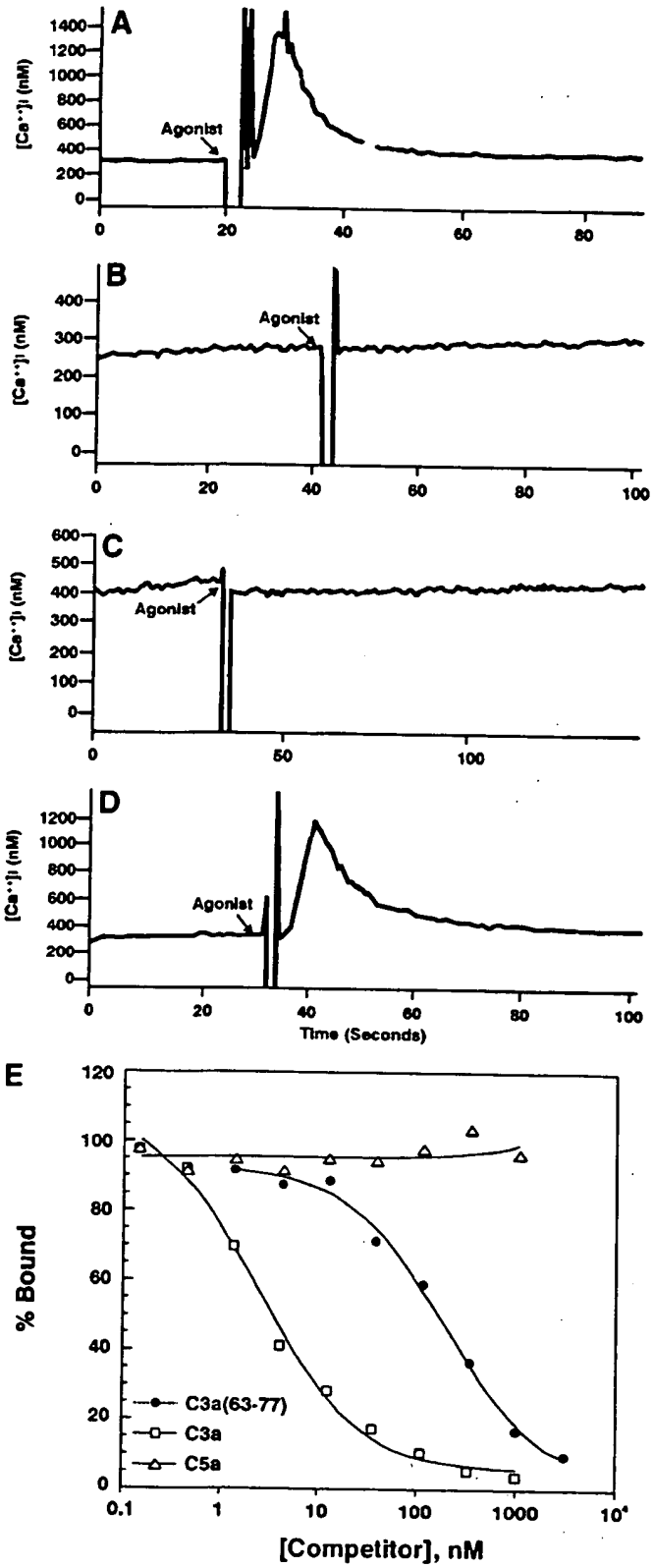


FIG. 3. **Cells expressing HNFAG09 but not C5a-R bind and respond to C3a.** Calcium mobilization by Fura 2-loaded cells expressing C5a-R (*A* and *C*) or HNFAG09 (*B* and *D*) in response to rC5a (10 nM (*A*) or 100 nM (*B*)) or C3a analogue peptide (1 μM, *C* and *D*) is shown. *E*, competition of ^{125}I-C3a binding to HNFAG09 expressing RBL-2H3 cells by increasing concentrations of C3a analogue synthetic peptide (*closed circle*), C3a (*open square*), or rC5a (*open triangle*).

HNFAG09-expressing cells responded to rC5a (Fig. 3, *A* and *B*). A robust response to a C3a carboxyl-terminal analogue synthetic peptide (WWGKKYRASKLGLAR) (9) (EC_{50} = 3.9 nM) was detected in cells expressing HNFAG09, but no response was obtained for C5a-R-expressing cells (Fig. 3, *D* and *C*, respectively). Similarly, HNFAG09 but not C5a-R expressing RBL-2H3 cells also responded to native human C3a (EC_{50} = 0.3 nM; data not shown).

C3a was radioiodinated and used in whole cell binding assays to further characterize HNFAG09. Binding of ^{125}I-C3a to HNFAG09 expressing RBL-2H3 cells was competed by increasing concentrations of C3a (IC_{50} = 3.0 nM) and the C3a analogue synthetic peptide (IC_{50} = 155 nM) but not by rC5a (Fig. 3*E*). By saturation binding and Scatchard analysis a single class of C3a binding sites was identified with an estimated K_d of 0.3 nM and a B_{max} of 32,000 receptors/cell (data not shown). Curiously, a HEK 293 cell line stably expressing HNFAG09 mRNA by Northern blot neither bound nor responded to C3a (data not shown).

RBL-2H3 cells expressing HNFAG09 bind and respond to C3a and a C3a analogue synthetic peptide but not C5a. These data, along with the results of the tissue distribution analysis, are consistent with HNFAG09 (AZ3B) (8) encoding the human C3a receptor.

The demonstration that C5a-R (reviewed in Ref. 20) and C3a-R expression is not limited to myeloid cells but that they both are expressed in a variety of non-myeloid cells throughout the body and that they are abundantly expressed in the central nervous system is consistent with these receptors having a much greater role in the pathogenesis of inflammatory and autoimmune diseases than previously suspected. Now that the receptor for C3a has been identified, further studies to elucidate the role of C3a in immune function and disease will be facilitated.

REFERENCES

1. Gerard, C., and Gerard, N. P. (1994) *Annu. Rev. Immunol.* **12**, 775–808
2. Hugli, T. E. (1984) *Springer Semin. Immunopathol.* **7**, 193–219
3. Bitter-Suermann, D. (1988) in *The Complement System* (Rother, K., and Till, G., eds) pp. 367–395, Springer-Verlag, Heidelberg
4. Vogt, W. (1986) *Complement* **3**, 177–188
5. Morgan, B. P. (1994) *Eur. J. Clin. Invest.* **24**, 219–228
6. Klos, A., Bank, S., Gietz, C., Bautsch, W., Köhl, J., Burg, M., and Kretzschmar, T. (1992) *Biochemistry* **31**, 11274–11282
7. Murakami, Y., Yamamoto, T., Imamichi, T., and Nagasawa, S. (1993) *Immunol. Lett.* **36**, 301–304
8. Roglic, A., Prossnitz, E. R., Cavanagh, S. L., Pan, Z., Zou, A., and Ye, R. D. (1996) *Biochim. Biophys. Acta* **1305**, 39–43
9. Ember, J. A., Johansen, N. L., and Hugli, T. E. (1991) *Biochemistry* **30**, 3603–3612
10. Jelinek, L. J., Lok, S., Rosenberg, G. B., Smith, R. A., Grant, F. J., Biggs, S., Bensch, P. A., Kuijper, J. L., Sheppard, P. O., Sprecher, C. A., O'Hara, P. J., Foster, D., Walker, K. M., Chen, L. H. J., McKernan, P. A., and Kindsvogel, W. (1993) *Science* **259**, 1614–1616
11. Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., and Venter, J. C. (1991) *Science* **252**, 1651–1656
12. Adams, M. D., Dubnick, M., Kerlavage, A. R., Moreno, R., Kelley, J. M., Kelly, J. M., Utterback, T. R., Nagle, J. W., Fields, C., and Venter, J. C. (1992) *Nature* **355**, 632–634
13. Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Buit, C. J., Lee, N. H., Kirkness, E. F., Weinstock, K. G., Gocayne, J. D., White, O., Sutton, G., Slake, J. A., Brandon, R. C., Chiu, M. W., Clayton, R. A., Cline, R. T., Cofton, M. D., Earle-Hughes, J., Fine, L. D., FitzGerald, L. M., FitzHugh, W. M., Fritchman, J. L., Geoghegan, N. S. M., Giodek, A., Gnehm, C. L., Hanna, M. C., Hodblom, E., Hinkle, P. S., Jr., Kelley, J. M., Klimek, K. M., Kelley, J. C., Liu, L-I., Marmaros, S. M., Merrick, J. M., Moreno-Palanques, R. F., McDonald, L. A., Nguyen, D. T., Pellegrino, S. M., Phillips, C. A., Ryder, S. E., Scott, J. L., Saudek, D. M., Shirley, R., Small, K. V., Spriggs, T. A., Ufterback, T. R., Weidman, J. F., Li, Y., Barthlow, R., Bednarik, D. P., Coo, L., Copeda, M. A., Coleman, T. A., Collins, E-J., Dimke, D., Fang, P., Ferris, A., Fischer, C., Hastings, G. A., He, W-W., Hu, J-S., Huddleston, K. A., Greene, J. M., Gruber, J., Hudson, P., Kim, A., Kozak, D. L., Kunsch, C., Hungjun, J., Li, H., Moissner, P. S., Olson, H., Raymond, L., Wei, Y-F., Wing, J., Xu, C., Yu, Q-L., Ruben, S. M., Dillon, P. J., Fannon, M. R., Rosen, C. A., Haseltine, W. A., Fields, C., Fraser, C. M., and Venter, J. C. (1995) *Nature* **377**, (suppl.) 3–174
14. Gerard, N. P., and Gerard, C. (1991) *Nature* **349**, 614–617

15. Boulay, F., Mery, L., Tardif, M., Brouchon, L., and Vignais, P. (1991) *Biochemistry* 30, 2993–2999
16. Aiyar, N., Baker, E., Wu, H-L., Nambi, P., Edwards, R. M., Trill, J. J., Ellis, C., and Bergsma, D. J. (1994) *Mol. Cell. Biol.* 131, 75–86
17. DeMartino, J. A., Van Riper, G., Siciliano, S. J., Molineaux, C. J., Konteatis, Z. D., Rosen, H., and Springer, M. S. (1994) *J. Biol. Chem.* 269, 14446–14450
18. Saussy, D. L., Jr., Sarau, H. M., Foley, J. J., Mong, S., and Crooke, S. T. (1989) *J. Biol. Chem.* 264, 19845–19855
19. Siraganian, R. P., McGivney, A., Barsumian, E. D., Crews, F. T., Hirata, F., and Axelrod, J. (1982) *Fed. Proc.* 41, 30–34
20. Wetsel, R. A. (1995) *Curr. Opin. Immunol.* 7, 48–53
21. Needleman, S. B., and Wunsch, C. D. (1970) *J. Mol. Biol.* 48, 443–453